

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: February 13, 2004, 08:23:26 ; Search time 9384 Seconds

(without alignments)  
11225.734 Million cell updates/sec

Title: US-09-836-613-1

Perfect score: 2575  
Sequence: 1 CCCCGGCTTACCTTCGGGT.....AGTCAGCGCGCCGGAATTC 2575

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_bcs:\*  
12: gb\_ey:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rod:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_ey:\*  
39: em\_hugo\_hum:\*  
40: em\_hugo\_mus:\*  
41: em\_hugo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	* Match	Length	DB	ID	Description
1	2575	100.0	2575	6	AR160672	AR160672 Sequence
2	2575	100.0	2575	9	HSU40846	U40846 Human alpha
3	2455	95.3	2458	9	HSU43573	U43573 Human alpha
4	2446	95.0	2446	9	HMDUPFSA	L78464 Human alpha
5	2063	80.1	2093	9	HMDUPFSA	L78465 Human alpha
6	1623	63.1	2452	10	MM085247	U85247 Mus musculus
7	1426	55.4	10127	9	HSU43572	U43572 Human alpha
8	1426	55.4	10380	6	AR160673	AR160673 Sequence
9	1426	55.4	46610	9	HSU34879	U34879 Human 17-be
10	1425	55.3	179146	9	AC067852	AC067852 Homo sapi
11	1415	55.0	114531	2	AC002537	AC002537 Homo sapi
12	1414	54.9	21764	9	HMD17BHD	M84472 Human 17-be
13	1114	43.3	2949	5	AF331668	AF331668 Dromaius
14	874	34.0	7721	10	AF003255	AF003255 Mus muscu
15	874	34.0	10938	10	AF363242	AF363242 Mus muscu
16	874	34.0	181557	10	AC069014	AC069014 Mus muscu
17	873	34.0	218666	2	EX323890	EX323890 Mus muscu
18	873	33.9	73944	10	EX255926	EX255926 Mus muscu
19	865	33.6	235178	2	AC098921	AC098921 Rattus no
20	687	26.7	5762	10	RN17BHD	X98038 R. norvegicu
21	584	22.7	4217	5	AF331669	AF331669 Dromaius
22	369	14.3	2759	3	AY058738	AY058738 Drosophila
23	369	14.3	36890	2	AC018157	AC018157 Drosophila
24	369	14.3	79076	3	AC004423	AC004423 Drosophila
25	369	14.3	186490	3	AC092224	AC092224 Drosophila
26	369	14.3	308031	3	AE003621	AE003621 Drosophila
27	286	11.1	2452	8	AY117179	AY117179 Arabidops
28	286	11.1	2702	8	AY080811	AY080811 Arabidops
29	258	10.0	2631	8	NFA18209	Y18209 Nicotiana t
30	209	8.1	10545	1	AE005727	AE005727 Caulobact
31	183	7.1	298550	1	AP005029	AP005029 Streptomy
32	179	7.0	9955	1	AE011700	AE011700 Xanthomon
33	172	6.7	30037	1	AE016927	AE016927 Bacteroid
34	170	6.6	30346	1	AE016941	AE016941 Bacteroid
35	155	6.0	299925	1	AP005045	AP005045 Streptomy
36	142	5.5	301761	1	AE016945	AE016945 Bacteroid
37	83	3.2	135552	2	AC019251	AC019251 Homo sapi
38	83	3.2	125020	2	AF429315	AF429315 Homo sapi
39	82	3.2	186752	9	AC111200	AC111200 Homo sapi
40	79	3.1	152686	2	AC143344	AC143344 Macaca mu
41	79	3.1	1094	11	PM7G11B	AL885196 Penicilliu
42	79	3.1	11305	8	OSAS35061	AJ535061 Oryza sat
43	79	3.1	159790	2	AP004786	AP004786 Oryza sat
44	79	3.1	65351	2	AC139773	AC139773 Homo sapi
45	79	3.1	69337	2	AC101261	AC101261 Mus muscu

## ALIGNMENTS

RESULT 1  
LOCUS AR160672 2575 bp DNA linear PAT 17-OCT-2001  
DEFINITION Sequence 1 from patent US 6255096.  
ACCESSION AR160672  
VERSION AR160672.1 GI:16225095  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2575)  
AUTHORS Hopwood, J., Joseph, J., Scott, H., Steele, J., Weber, B., Blanch, L. and Anson, D., Stewart.  
TITLE Synthetic mammalian .alpha.-n-acetylglucosaminidase and genetic sequences encoding same

JOURNAL Patent: US 6255096-A 1 03-JUL-2001;  
 FEATURES Location/Qualifiers  
 source 1..2575  
 /organism="unknown"  
 BASE COUNT 440 a 819 c 833 g 483 t  
 ORIGIN

Query Match 100.0%; Score 2575; DB 6; Length 2575;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CCCGGGCTTAAAGCTTGGGGTCCACGTGCGCGGAGCCCGAGCTGATTGACCCCGGCGG 60
Db 1 CCCGGGCTTAAAGCTTGGGGTCCACGTGCGCGGAGCCCGAGCTGATTGACCCCGGCGG 60
QY 61 CCCGACCCCTTGCGCGCTGCGGAGCCCGCAGAGACTGAGACCATGAGAGCGGTGAGTGG 120
Db 61 CCCGACCCCTTGCGCGCTGCGGAGCCCGCAGAGACTGAGACCATGAGAGCGGTGAGTGG 120
QY 121 CCGCGGCGGTGGGGGTCTCTCTCTGAGCCGGGGCCGGGAGCGGAGAGAGAGAGAGAGAG 180
Db 121 CCGCGGCGGTGGGGGTCTCTCTCTGAGCCGGGGCCGGGAGCGGAGAGAGAGAGAGAGAG 180
QY 181 GGGAGGCGGCGGCGGTGCGGCGGCTGCTGAGCCGCGCTGCTGAGGCGGAGGCGGCGG 240
Db 181 GGGAGGCGGCGGCGGTGCGGCGGCTGCTGAGCCGCGCTGCTGAGGCGGAGGCGGCGG 240
QY 241 ACTTCTCCGTGTGCTGCTGAGAGCGCTTGTGCTGCGCAAGCCGAGCTTGGACACTTAC 300
Db 241 ACTTCTCCGTGTGCTGCTGAGAGCGCTTGTGCTGCGCAAGCCGAGCTTGGACACTTAC 300
QY 301 TGGGCGGCGGCGGCGGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCGGCGGCG 360
Db 301 TGGGCGGCGGCGGCGGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCGGCGGCG 360
QY 361 CCGGGCTGCAACCGCTACCTGCGGACTTGTGTGCTGCGCAAGCTGAGCTGAGCTGCTC 420
Db 361 CCGGGCTGCAACCGCTACCTGCGGACTTGTGTGCTGCGCAAGCTGAGCTGAGCTGCTC 420
QY 421 AGCTGCGGCTGCGGCGGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCGGCGG 480
Db 421 AGCTGCGGCTGCGGCGGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCGGCGG 480
QY 481 ACAAGTACCGCTATTACAGAAATGTGTGCAAGCAAACTACTCTGCTGTGTGTGAGT 540
Db 481 ACAAGTACCGCTATTACAGAAATGTGTGCAAGCAAACTACTCTGCTGTGTGTGAGT 540
QY 541 GGGGCGGCTGCGGAGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Db 541 GGGGCGGCTGCGGAGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 601 CTTGAGAGCGGCGAGAGAGCGGCTGAGGAGGAGTACCTGAGCTTGGGCGTGAACCCAG 660
Db 601 CTTGAGAGCGGCGAGAGAGCGGCTGAGGAGGAGTACCTGAGCTTGGGCGTGAACCCAG 660
QY 661 CAGAGATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
Db 661 CAGAGATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 721 TGCACACCTGAGGATGAGGCGGCGGCGGCGGCTGAGGAGTCAAGAGAGCTTAACTGAC 780
Db 721 TGCACACCTGAGGATGAGGCGGCGGCGGCGGCTGAGGAGTCAAGAGAGCTTAACTGAC 780
QY 781 ACCGGGCTCTGAGACAGAGTGCCTCTTGGGAGTGAACCCAGTGCCTGATTTGGCGG 840
Db 781 ACCGGGCTCTGAGACAGAGTGCCTCTTGGGAGTGAACCCAGTGCCTGATTTGGCGG 840
QY 841 GGGATGTTCCCGAGGCTGTGACACAGAGGTTCCCTCAGGTCAATGTCAAGAAATGGGGA 900
Db 841 GGGATGTTCCCGAGGCTGTGACACAGAGGTTCCCTCAGGTCAATGTCAAGAAATGGGGA 900
QY 901 GTTGGGCGCACTTAACTGTCTACTCTGCTCTTCTCTTCTGAGCTTCCGAGAGACCCCA 960
  
```

```

Db 901 GTTGGGCGCACTTAACTGTCTACTCTGCTCTTCTCTTCTGAGCTTCCGAGAGACCCCA 960
QY 961 TATTCCTCATGATGGGAGCGCTTCCGAGAGAGCTGATCAAGAGATTGGGACAGAGC 1020
Db 961 TATTCCTCATGATGGGAGCGCTTCCGAGAGAGCTGATCAAGAGATTGGGACAGAGC 1020
QY 1021 ACATCTATGAGGCGGACACTTTCATATGATGACAGCCACTTCTCTCAGAGCCCTCTTACC 1080
Db 1021 ACATCTATGAGGCGGACACTTTCATATGATGACAGCCACTTCTCTCAGAGCCCTCTTACC 1080
QY 1081 TTGCGGAGCCACACATGCGGCTATGAGGCGATGACATGACATGATGATGATGATGATGAT 1140
Db 1081 TTGCGGAGCCACACATGCGGCTATGAGGCGATGACATGACATGATGATGATGATGATGAT 1140
QY 1141 GAGTCTCTCAAGAGCTGCTCTTCCAGACACAGCCGAGTTCGGGGGCGGCGCCAGATCA 1200
Db 1141 GAGTCTCTCAAGAGCTGCTCTTCCAGACACAGCCGAGTTCGGGGGCGGCGCCAGATCA 1200
QY 1201 GGGCTGTGCTGAGAGCTGTGCTCCGTGCGGCTCTCTGCTTCTGAGACTGTTTGTGAGA 1260
Db 1201 GGGCTGTGCTGAGAGCTGTGCTCCGTGCGGCTCTCTGCTTCTGAGACTGTTTGTGAGA 1260
QY 1261 GCGAGCTGTGTATACCGGCACTGCTCTTCCAGAGCCAGCCCTTATCTGATCTGTGATGC 1320
Db 1261 GCGAGCTGTGTATACCGGCACTGCTCTTCCAGAGCCAGCCCTTATCTGATCTGTGATGC 1320
QY 1321 TGCACAACTTTGGGGGAAACCATGCTCTTTTGGAGCCCTTAGAGGCTGTGAAACGAGGCC 1380
Db 1321 TGCACAACTTTGGGGGAAACCATGCTCTTTTGGAGCCCTTAGAGGCTGTGAAACGAGGCC 1380
QY 1381 CAGAGGCTGCGGCTCTTCCCACTCAACCATGATGAGGACGGGATGAGCCCGGAGG 1440
Db 1381 CAGAGGCTGCGGCTCTTCCCACTCAACCATGATGAGGACGGGATGAGCCCGGAGG 1440
QY 1441 GCATCAGCAGAAAGAAATGAGTATTCCTCATGAGCTGAGCTGAGGAGAAAGAGC 1500
Db 1441 GCATCAGCAGAAAGAAATGAGTATTCCTCATGAGCTGAGCTGAGGAGAAAGAGC 1500
QY 1501 CAGTGCAGATTTGAGGAGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
Db 1501 CAGTGCAGATTTGAGGAGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
QY 1561 ACCCGAGCAGAGGAGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
Db 1561 ACCCGAGCAGAGGAGCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
QY 1621 AGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
Db 1621 AGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
QY 1681 CCAAGATCTGATCAACCGATCTGATGATGATGAGGCTGAGGAGGAGGAGGAGGAGGAG 1740
Db 1681 CCAAGATCTGATCAACCGATCTGATGATGATGAGGCTGAGGAGGAGGAGGAGGAGGAG 1740
QY 1741 CTCCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800
Db 1741 CTCCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800
QY 1801 CAGTGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1860
Db 1801 CAGTGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1860
QY 1861 AGTGGGCTCTCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
Db 1861 AGTGGGCTCTCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
QY 1921 ACGAGGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
Db 1921 ACGAGGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
QY 1981 CAGGCGAGTCAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
Db 1981 CAGGCGAGTCAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
  
```

QY 2041 CCTTGTGGGGGCGAAGGCAACATCTGGACTATGCAACAAGCAGCTGGCGGGTTGG 2100  
 Db 2041 CCTTGTGGGGGCGAAGGCAACATCTGGACTATGCAACAAGCAGCTGGCGGGTTGG 2100  
 QY 2101 TGGCCAACTCTACACCCCTGGCTGGGGCTTTTCTGGAGGCGCTGGTTGACAGTGTGG 2160  
 Db 2101 TGGCCAACTCTACACCCCTGGCTGGGGCTTTTCTGGAGGCGCTGGTTGACAGTGTGG 2160  
 QY 2161 CCCAGGCGATCCCTTTCACACAGCAACAGTTTGACAAAAATGTCTTCAACTGGAGCAGG 2220  
 Db 2161 CCCAGGCGATCCCTTTCACACAGCAACAGTTTGACAAAAATGTCTTCAACTGGAGCAGG 2220  
 QY 2221 CCTTGTCTGACGACAGAGGTAACCCAGCCAGCCGCGAGGAGACACTGTGGACTGG 2280  
 Db 2221 CCTTGTCTGACGACAGAGGTAACCCAGCCAGCCGCGAGGAGACACTGTGGACTGG 2280  
 QY 2281 CCAAGAAAGATCTTCCCAATATTTACCCCGGGCTGGGGTGGCCGCTGTGTGATAGATTG 2340  
 Db 2281 CCAAGAAAGATCTTCCCAATATTTACCCCGGGCTGGGGTGGCCGCTGTGTGATAGATTG 2340  
 QY 2341 GCCACCACTGGGCTTGTCTTTCGCTAATTCAGGGCAGATTTCAGAGGCCAGAGCTGGA 2400  
 Db 2341 GCCACCACTGGGCTTGTCTTTCGCTAATTCAGGGCAGATTTCAGAGGCCAGAGCTGGA 2400  
 QY 2401 CAGACATCAAGAGATTAACCCAGGCTGGAGAGGCGCCCAAGGCTGTGTGGGTCTG 2460  
 Db 2401 CAGACATCAAGAGATTAACCCAGGCTGGAGAGGCGCCCAAGGCTGTGTGGGTCTG 2460  
 QY 2461 ACCTGGGGGAGTTGAGGAGGAAATGACCTGCTCCACACACCCAAAGTGTGGATTGA 2520  
 Db 2461 ACCTGGGGGAGTTGAGGAGGAAATGACCTGCTCCACACACCCAAAGTGTGGATTGA 2520  
 QY 2521 AGTACTGTTTCTTCTTCACTTAATAAAAAAGTGGAGGCGCGGAATTC 2575  
 Db 2521 AGTACTGTTTCTTCTTCACTTAATAAAAAAGTGGAGGCGCGGAATTC 2575

RESULT 2  
 HSU40846 2575 bp mRNA linear PRI 15-AUG-1996  
 LOCUS HSU40846 2575 bp mRNA linear PRI 15-AUG-1996  
 DEFINITION Human alpha-N-acetylglucosaminidase (NAG) mRNA, complete cds.  
 VERSION U40846  
 ACCESSION U40846.1 GI:1197839  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 2575)  
 AUTHORS Weber, B., Blanch, L., Clements, P.R., Scott, H.S. and Hopwood, J.J.  
 TITLE Cloning and expression of the gene involved in Sanfilippo B syndrome (mucopolysaccharidosis III B)  
 JOURNAL Hum. Mol. Genet. 5 (6), 771-777 (1996)  
 MEDLINE 96372812  
 PUBMED 8776591  
 REFERENCE 2 (bases 1 to 2575)  
 AUTHORS Weber, B., Scott, H. and Hopwood, J.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-NOV-1995) Birgit Weber, Chemical Pathology, Women's & Children's Hospital, King William Street 72, North Adelaide, S.A. 5006, Australia  
 FEATURES  
 source  
 1. .2575  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="11q12-21"  
 1. .2575  
 /gene="NAG"  
 102. .2333  
 /gene="NAG"

BASE COUNT 440 a 819 c 833 g 483 t  
 ORIGIN  
 Query Match 100.0%; Score 2575; DB 9; Length 2575;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 CCCGGGCTTACCTTGGGCTCACGTGGCCGAGAGCCGCGAGCTGATTGACGCGGCGG 60  
 1 CCCGGGCTTACCTTGGGCTCACGTGGCCGAGAGCCGCGAGCTGATTGACGCGGCGG 60  
 61 CCCACCCCTGGGCGCGTCCGCGGAGCCGCGAGAGTGAACCATGAGAGCGGTGGCGGTG 120  
 61 CCCACCCCTGGGCGCGTCCGCGGAGCCGCGAGAGTGAACCATGAGAGCGGTGGCGGTG 120  
 61 CCCACCCCTGGGCGCGTCCGCGGAGCCGCGAGAGTGAACCATGAGAGCGGTGGCGGTG 120  
 121 CCGCGCGGTGGGGGTCTTCTCTCTGGCCGCGGCGCGGCGCGCGAGCGAGCGAGCGCC 180  
 121 CCGCGCGGTGGGGGTCTTCTCTCTGGCCGCGGCGCGGCGCGCGAGCGAGCGAGCGCC 180  
 121 CCGCGCGGTGGGGGTCTTCTCTCTGGCCGCGGCGCGGCGCGCGAGCGAGCGAGCGCC 180  
 181 GCGAGCGCGCGCGCTGCGGCGCGCTGCGTGGCCGCGGCTGCGGCGCGAGCGCGCGCG 240  
 181 GCGAGCGCGCGCGCTGCGGCGCGCTGCGTGGCCGCGGCTGCGGCGCGAGCGCGCGCG 240  
 241 ACTTCTCGGTGCGGTGGAGCGCTGCTGGCTGCGCAAGCGCGCTTGAACAACCTACAGCC 300  
 241 ACTTCTCGGTGCGGTGGAGCGCTGCTGGCTGCGCAAGCGCGCTTGAACAACCTACAGCC 300  
 241 ACTTCTCGGTGCGGTGGAGCGCTGCTGGCTGCGCAAGCGCGCTTGAACAACCTACAGCC 300  
 301 TGGGCGGCGGCGCGCGCGCGCTGCGGCTGCGGCTGCGGCGCTGCGGCGCGCGCGCG 360  
 301 TGGGCGGCGGCGCGCGCGCGCTGCGGCTGCGGCTGCGGCGCTGCGGCGCGCGCGCGCG 360  
 361 CGGGGCTGCAACCGCTTACCTGCGGCACTTCTGTGTGCTGCCAGGTGGCCTGTGTGCGGCTCTC 420  
 361 CGGGGCTGCAACCGCTTACCTGCGGCACTTCTGTGTGCTGCCAGGTGGCCTGTGTGCGGCTCTC 420  
 421 AGTGGCGCTTGGCGCGCGCACTGCGAGCCGTGCGGCGGAGCTGACCGAGCGCACGCGCA 480  
 421 AGTGGCGCTTGGCGCGCGCACTGCGAGCCGTGCGGCGGAGCTGACCGAGCGCACGCGCA 480  
 421 AGTGGCGCTTGGCGCGCGCACTGCGAGCCGTGCGGCGGAGCTGACCGAGCGCACGCGCA 480  
 481 ACAGGTACCGGTATTACAGAGATGTGGACGCGAAGAACTCTCTTGTGTGTGTGTGTGTGT 540  
 481 ACAGGTACCGGTATTACAGAGATGTGGACGCGAAGAACTCTCTTGTGTGTGTGTGTGTGT 540  
 481 ACAGGTACCGGTATTACAGAGATGTGGACGCGAAGAACTCTCTTGTGTGTGTGTGTGTGT 540  
 541 GGGCGCGCTGGAGCGAGAGATGAGTGGCTGGAATGGCATGAACCTGGCACTGG 600  
 541 GGGCGCGCTGGAGCGAGAGATGAGTGGCTGGAATGGCATGAACCTGGCACTGG 600  
 541 GGGCGCGCTGGAGCGAGAGATGAGTGGCTGGAATGGCATGAACCTGGCACTGG 600  
 601 CTTGAGGCGCGAGAGAGCATCTGGAGCGGGGTGACTGGCTTGTGGCTTGAACCGAGG 660  
 601 CTTGAGGCGCGAGAGAGCATCTGGAGCGGGGTGACTGGCTTGTGGCTTGAACCGAGG 660  
 661 CAGAGATCAATGAGTCTTACTGTGCTGCTTCTGCGCGCTGGGGCGAATGGGCAACC 720

Db	661	CAGAGATCAATGAGTTCCTTA	CTGAGTCCTG	CTCCTCTG	GGCGGGAATG	GGGCAAC	720			
QY	721	TGCA	CACCTGGAGATG	CCCCCTG	CCCCCTCT	TGGCAATCA	GAGTCTTA	CTGCA	GC	780
Db	721	TGCA	CACCTGGAGATG	CCCCCTG	CCCCCTCT	TGGCAATCA	GAGTCTTA	CTGCA	GC	780
QY	781	ACC	GGTCTTGAGACAGATG	GGCTCTT	GGCATAG	CCCCAGTGTG	CTCATTC	GGG		840
Db	781	ACC	GGGCTCTG	AGACAGATG	GGCTCTT	GGCATAG	CCCCAGTGTG	CTCATTC	GGG	840
QY	841	GGA	TGTTCCG	AGGCTGTCA	CA	GGGGTGTTCCT	CAAGTCA	TGTCA	CGAAGATG	900
Db	841	GGA	TGTTCCG	AGGCTGTCA	CA	GGGGTGTTCCT	CAAGTCA	TGTCA	CGAAGATG	900
QY	901	GTT	GGGGCCAC	TTTACTG	TCTACT	CTCTGCTCTT	CTTCTG	CTCGGAAG	CCCCCA	960
Db	901	GTT	GGGGCCAC	TTTACTG	TCTACT	CTCTGCTCTT	CTTCTG	CTCGGAAG	CCCCCA	960
QY	961	TAT	TCCCAT	CATCGG	AGCCTCTT	CTGCA	GAGCTGAT	CA	MAAGTTT	1020
Db	961	TAT	TCCCAT	CATCGG	AGCCTCTT	CTGCA	GAGCTGAT	CA	MAAGTTT	1020
QY	1021	ACAT	TATGGGGCCG	CACTTGCAT	TGAGATG	CA	CCACTTCT	CTCA	AGCCCTCT	1080
Db	1021	ACAT	TATGGGGCCG	CACTTGCAT	TGAGATG	CA	CCACTTCT	CTCA	AGCCCTCT	1080
QY	1081	TTG	CCGACAG	CCACCACTG	CCGCTAT	TGAG	CCCATG	ATGAGTGA	TACTAG	1140
Db	1081	TTG	CCGACAG	CCACCACTG	CCGCTAT	TGAG	CCCATG	ATGAGTGA	TACTAG	1140
QY	1141	GCG	TGCTC	CA	AGGCTG	CTTTC	CA	GCAC	CA	1200
Db	1141	GCG	TGCTC	CA	AGGCTG	CTTTC	CA	GCAC	CA	1200
QY	1201	GAG	CTGCTG	GGAGCTG	TG	CCGCGGCTG	CTGAT	CGA	CCCTG	1260
Db	1201	GAG	CTGCTG	GGAGCTG	TG	CCGCGGCTG	CTGAT	CGA	CCCTG	1260
QY	1261	GCG	AGCTGT	GTAT	ATAC	CCGCA	CTGCTCTT	CA	GGGCA	1320
Db	1261	GCG	AGCTGT	GTAT	ATAC	CCGCA	CTGCTCTT	CA	GGGCA	1320
QY	1321	TGCA	CACTTTGGGGG	AAACAT	TGCTCTTTT	TGAG	CCCTG	AGGCTG	TGAACG	1380
Db	1321	TGCA	CACTTTGGGGG	AAACAT	TGCTCTTTT	TGAG	CCCTG	AGGCTG	TGAACG	1380
QY	1381	CAGA	AGCTG	CCGCTCTT	TCCCA	CTCCAC	CATG	TATG	AGGCA	1440
Db	1381	CAGA	AGCTG	CCGCTCTT	TCCCA	CTCCAC	CATG	TATG	AGGCA	1440
QY	1441	GCA	TAGCC	CA	GAAGT	GTCTAT	TTCCT	CTAT	GGCTG	1500
Db	1441	GCA	TAGCC	CA	GAAGT	GTCTAT	TTCCT	CTAT	GGCTG	1500
QY	1501	CAG	TCCAG	ATTTGG	CAGCT	TGGAG	CAAGTTT	GC	CGCGG	1560
Db	1501	CAG	TCCAG	ATTTGG	CAGCT	TGGAG	CAAGTTT	GC	CGCGG	1560
QY	1561	ACC	CGAG	CGACAG	GGCTG	TGAG	AGTCT	CCGAG	TGTCA	1620
Db	1561	ACC	CGAG	CGACAG	GGCTG	TGAG	AGTCT	CCGAG	TGTCA	1620
QY	1621	AGG	CTG	AGAGG	CCACAT	GT	TACCCG	CTG	AGGCG	1680
Db	1621	AGG	CTG	AGAGG	CCACAT	GT	TACCCG	CTG	AGGCG	1680
QY	1681	CCAG	CAT	TG	GTACA	CCGAT	CTGA	TGT	TGAG	1740
Db	1681	CCAG	CAT	TG	GTACA	CCGAT	CTGA	TGT	TGAG	1740
QY	1741	CTC	CT	CT	CT	CT	CT	CT	CT	1800
Db	1741	CTC	CT	CT	CT	CT	CT	CT	CT	1800

QY	1801	CAGGCGACGAGACTGCTGAGCTTGACTACTATATAGAGAGGCAAGAAAGGCGCTTACCTGACCAAG	1860
Db	1801	CAGTGCAGAGACTGCTGAGCTTGACTACTATATAGAGAGGCAAGAAAGGCGCTTACCTGACCAAG	1860
QY	1861	AGTGTGCTCTCCCTGTTAGAGGACTGAGAGCGCTCCTGTATGAGCTGTGCTGCGCACTGG	1920
Db	1861	AGCTGTGCTCTCCCTGTTAGAGGACTGAGAGCGCTCCTGTATGAGCTGTGCTGCGCACTGG	1920
QY	1921	ACGAGGTGTGTGTGCTGAGTGAACAGCCGCTTCTTGCTGTGGCAGCTGGCTGAGAGAGGCCGAG	1980
Db	1921	ACGAGGTGTGTGTGCTGAGTGAACAGCCGCTTCTTGCTGTGGCAGCTGGCTGAGAGAGGCCGAG	1980
QY	1981	CAGCGGACGTCAAGTGAAGCGCGAGGCGCGATTTCTACGAGCAAGACGCGCTACCAAGTGA	2040
Db	1981	CAGCGGACGTCAAGTGAAGCGCGAGGCGCGATTTCTACGAGCAAGACGCGCTACCAAGTGA	2040
QY	2041	CGTTGTGGGGGGCCAGAAAGGCAACACTCTGTGACTATGCAACAAAGCACTGGCGGGGGTGG	2100
Db	2041	CGTTGTGGGGGGCCAGAAAGGCAACACTCTGTGACTATGCAACAAAGCACTGGCGGGGGTGG	2100
QY	2101	TGGCCAACTACTACACCCCTCGCTGCGGCGCTTTCTCGAAGCGCGCTGTTGACAGTGTG	2160
Db	2101	TGGCCAACTACTACACCCCTCGCTGCGGCGCTTTCTCGAAGCGCGCTGTTGACAGTGTG	2160
QY	2161	CCGAGGGCATTCCTCTTCACACAGACCAAGTATTGACAAAAATGCTTCCACTGTGAGCAGG	2220
Db	2161	CCGAGGGCATTCCTCTTCACACAGACCAAGTATTGACAAAAATGCTTCCACTGTGAGCAGG	2220
QY	2221	CGTTGTGTTCTCAGCAACAGAGGTATCCCGACCGCGCGAGAGACACTGTGTGACCTGG	2280
Db	2221	CGTTGTGTTCTCAGCAACAGAGGTATCCCGACCGCGCGAGAGACACTGTGTGACCTGG	2280
QY	2281	CCAAAGAAAGATCTTCTCAATATTTACCCCGGCTGTGGTGGCCGCTCTTGCTGATATGATTC	2340
Db	2281	CCAAAGAAAGATCTTCTCAATATTTACCCCGGCTGTGGTGGCCGCTCTTGCTGATATGATTC	2340
QY	2341	GCCACCACTGGGCGCTTGTTCGCTAATTCACAGGGCAGATTCCAGGGGCCACAGAGCTGGA	2400
Db	2341	GCCACCACTGGGCGCTTGTTCGCTAATTCACAGGGCAGATTCCAGGGGCCACAGAGCTGGA	2400
QY	2401	CAGACATCACAGGATTAACCCAGGCTGTGGAGAGAGGCCCCACGCGCTCTGTGTGGGGTCTG	2460
Db	2401	CAGACATCACAGGATTAACCCAGGCTGTGGAGAGAGGCCCCACGCGCTCTGTGTGGGGTCTG	2460
QY	2461	ACCGGGGGGGGATTGGAGGGAAATGACTGCGCCTCCACCAACCAAGTGTGGGATTTA	2520
Db	2461	ACCGGGGGGGGATTGGAGGGAAATGACTGCGCCTCCACCAACCAAGTGTGGGATTTA	2520
QY	2521	AGTACTGTCTTTCTTTCCACTTTAAAAAAGTGTGAGCGGCGCGGATTC	2575
Db	2521	AGTACTGTCTTTCTTTCCACTTTAAAAAAGTGTGAGCGGCGCGGATTC	2575

**RESULT 3**  
**HSU43573**  
**LOCUS** HSU43573 2458 bp mRNA linear PRI 11-JUN-1996  
**DEFINITION** Human alpha-N-acetylglucosaminidase (NAGLU) mRNA, complete cds.  
**ACCESSION** U43573  
**VERSION** U43573.1 GI:1171230  
**KEYWORDS**  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
**REFERENCE**  
**AUTHORS** Zhao,H.G., Li,H.H., Bach,G., Schmidtchen,A. and Neufeld,E.F.  
**TITLE** The molecular basis of Sanfilippo syndrome type B  
**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 93 (12), 6105-6105 (1996)  
**PMID** 8634097  
**PMED** 8650226  
**REFERENCE** 2 (bases 1 to 2458)  
**AUTHORS** Zhao,H.G., Li,H.H. and Neufeld,E.F.

TITLE Direct Submission  
 JOURNAL Submitted (19-DEC-1995) Hong G. Zhao, Biological Chemistry, UCLA,  
 10833 Le Conte Ave., Los Angeles, CA 90095, USA  
 FEATURES Location/Qualifiers

source  
 1. .2458  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17q21"  
 /tissue type="testis"  
 1. .2458  
 /gene="NAGLU"  
 1. .2232  
 /gene="NAGLU"  
 /EC\_number="3.2.1.50"  
 /function="one of four enzymes involved in the degradation  
 of heparan sulfate; specifically removes the  
 alpha-N-acetylglucosamine residues"  
 /note="Lysosomal enzyme; deficient in Sanfilippo B  
 syndrome"  
 /codon\_start=1  
 /product="alpha-N-acetylglucosaminidase"  
 /protein\_id="AAC50513.1"  
 /db\_xref="GI:1171231"  
 /translation="MEAVAVAAVGLLHAGAGAGDEAREAAVRAVRLGPGP  
 AAPSVSRVLAIAKPLDTSLSGGGAAVRVSGTGYAAAGAHYLRDCCGHVA  
 WSSQJLRPRPLPVPGLTEATPNRYQNVCTOSYFTMDMAREREDIMALN  
 HINLALMSQGEAIWQRYVIALGLTQALINEFFTGPAFLWGMKGNLWDPLPSW  
 HIQYLAHQHVDQMSFGMTPLYPAFAHVEPAVRVPOVNTKMSWGFNCYS  
 CSFLAPDEPIPIIGSLFRELKEFGTDHLYGADTFEMQPPSEBPSYLAATAV  
 YEAMTADTEAVMLQGLFQHPQFQPAQIRAVGVAPRRLVLDLFAASQVYT  
 RTASFOGQPIWCMHNFQNGHGLFALANVNGPEAAFLPNSITVWGMAPEGISO  
 NRVYSIMEIEMKRCQVPDLAAWTSPAPRYGVSHPRAGAAWLLLSVYNSGEA  
 CRHANSPLVRPSLQMTSIVYKNSDYFAHMLLTSPLATSPAFYDLDDTRO  
 AVEGLVSLYEBARSAYLSKELASLIRAGVLAIEILLPADVLASDSRYLLGSWQ  
 ARAAASLEADAFYEQNSRYQLTLMGPENLIDYANKQLAGLVANYTPRMFLFLEAL  
 VDSVAQGIPIFOHQHFDKNVFOLEQAFVLSKORVPSQPRGDTVDLAKKIFLKYPGMVA  
 GSW"

BASE COUNT 428 a 773 c 789 g 468 t

Query Match 95.3%; Score 2455; DB 9; Length 2458;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 ATGGAGCGGTGGCGGTGGCGCGGGGCTCTCTGCGCGGGGCGGGGGG 161  
 Db 1 ATGGAGCGGTGGCGGTGGCGCGGGGCTCTCTCTGCGCGGGGCGGGGGG 60  
 QY 162 GCGGCAAGGCGAGAGGCGCGGCGCGCGCGCTCGTGACCGCGCTGTG 221  
 Db 61 GCGGCAAGGCGAGAGGCGCGGCGCGCGCGCTCGTGACCGCGCTGTG 120  
 QY 222 GGGCGAGCGCGCGCGCGCGCGCTCTCTCTGCGCGGGGCGGGGGG 281  
 Db 121 GGGCGAGCGCGCGCGCGCGCTCTCTCTGCGCGGGGCGGGGGG 180  
 QY 282 GGGCTGAGCACTTACAGCTGCGCGCGCGCGCGCGCTGCGGGTGC 341  
 Db 181 GGGCTGAGCACTTACAGCTGCGCGCGCGCGCGCGCTGCGGGTGC 240  
 QY 342 ACGGCGCTGGCGCGCGCGCGCGCGCTTCTCTGCGCGCGCGCGCG 401  
 Db 241 ACGGCGCTGGCGCGCGCGCGCGCGCTTCTCTGCGCGCGCGCGCG 300  
 QY 402 GTGCGCTGTCCGAGCTCTGAGCTGCGCGCGCGCGCGCTGCGCGG 461  
 Db 301 GTGCGCTGTCCGAGCTCTGAGCTGCGCGCGCGCGCGCTGCGCGG 360  
 QY 462 CTGACCGAGGCGACGCCCAAGGTACCGTATTACCAAGATGTGCA 521  
 Db 361 CTGACCGAGGCGACGCCCAAGGTACCGTATTACCAAGATGTGCA 420

QY 522 TCCTTCGTGTGTGAGACTGGGCGCGCGCTGGAGCGAGAGATAGCTGATGCGCTAAT 581  
 Db 421 TCCTTCGTGTGTGAGACTGGGCGCGCGCTGGAGCGAGAGATAGCTGATGCGCTAAT 480  
 QY 582 GGCATCAACCTGGGCACTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 641  
 Db 481 GGCATCAACCTGGGCACTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
 QY 642 GCGTTGGGCGTGAACCGAGGCGAGAGATCAATGAGTCTTTTACTGTGCTGCTGCGCG 701  
 Db 541 GCGTTGGGCGTGAACCGAGGCGAGAGATCAATGAGTCTTTTACTGTGCTGCTGCGCG 600  
 QY 702 TGGGGGCGAATGGGCGAAGCTGCAACCTGGGATGGGCGCGCGCGCGCGCGCGCGCG 761  
 Db 601 TGGGGGCGAATGGGCGAAGCTGCAACCTGGGATGGGCGCGCGCGCGCGCGCGCGCG 660  
 QY 762 AAGCAGCTTTTACCTGACGACCGGCGTCTGGACCGAGATGCGCTCTTGGCANTGACCCA 821  
 Db 661 AAGCAGCTTTTACCTGACGACCGGCGTCTGGACCGAGATGCGCTCTTGGCANTGACCCA 720  
 QY 822 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 881  
 Db 721 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
 QY 882 AATGTCAGAAAGATGGGCGAGTTGGGCGCACTTTTACTGTTCTTCTGCTGCTGCTGCT 941  
 Db 781 AATGTCAGAAAGATGGGCGAGTTGGGCGCACTTTTACTGTTCTTCTGCTGCTGCTGCT 840  
 QY 942 CTGGCTCGGAAGACCCCATATTTCCCATCATCGGAGGCTCTTCTGCGAGAGCTATATC 1001  
 Db 841 CTGGCTCGGAAGACCCCATATTTCCCATCATCGGAGGCTCTTCTGCGAGAGCTATATC 900  
 QY 1002 AAAGAGTTGGCAAGACCAATATATATGAGGCGGCGCACTTTCAATGAGATGACGCCACT 1061  
 Db 901 AAAGAGTTGGCAAGACCAATATATATGAGGCGGCGCACTTTCAATGAGATGACGCCACT 960  
 QY 1062 TCCTCAGAGCCCTCTTACTGTCGCGAGCCACCACTGCGCTTATGAGAGCCATGATGCA 1121  
 Db 961 TCCTCAGAGCCCTCTTACTGTCGCGAGCCACCACTGCGCTTATGAGAGCCATGATGCA 1020  
 QY 1122 GTGATATATGAGGCTGT 1181  
 Db 1021 GTGATATATGAGGCTGT 1080  
 QY 1182 TGGGCGCGCGCGCGCGCGCGCTGTGCTGCGAGCTGTGCGCGCGCGCGCGCTGCTGCT 1241  
 Db 1081 TGGGCGCGCGCGCGCGCGCGCTGTGCTGCGAGCTGTGCGCGCGCGCGCGCTGCTGCT 1140  
 QY 1242 CTGACCTGTTTGTGTGAGAGCGAGCTGTGTATACCGGCACTGCTCTTTCAGAGGCGAG 1301  
 Db 1141 CTGACCTGTTTGTGTGAGAGCGAGCTGTGTATACCGGCACTGCTCTTTCAGAGGCGAG 1200  
 QY 1302 CCGTTCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1361  
 Db 1201 CCGTTCAATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1260  
 QY 1362 GAGGCTGTGAACGAGGCGCGAGGCTGTGCGCGCGCTTTCGCAACCTGACCATGTGATG 1421  
 Db 1261 GAGGCTGTGAACGAGGCGCGAGGCTGTGCGCGCGCTTTCGCAACCTGACCATGTGATG 1320  
 QY 1422 ACGGCGATGCGCGCGCGCGCGCGAGGATGAGCGAGGAGGATGAGGATGAGGATGAGG 1481  
 Db 1321 ACGGCGATGCGCGCGCGCGCGCGAGGATGAGCGAGGAGGATGAGGATGAGGATGAGG 1380  
 QY 1482 CTGGGCTGGCGAAGAGACCGAGTGCAGATTTGGCAGCTTGGGATGACAGCTTTTGGCGCC 1541  
 Db 1381 CTGGGCTGGCGAAGAGACCGAGTGCAGATTTGGCAGCTTGGGATGACAGCTTTTGGCGCC 1440  
 QY 1542 CGGCGGATGAGGAGTCTCCACCCGAGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGG 1601  
 Db 1441 CGGCGGATGAGGAGTCTCCACCCGAGCGAGGCGAGGCGAGGCGAGGCGAGGCGAGG 1500

QY 1602 GTGTCACTGCTCCGGGAGGCTTGACAGGGCCCAATCTAGCCCGCTGCTACAGGG 1661  
 DB 1501 GTGTCACTGCTCCGGGAGGCTTGACAGGGCCCAATCTAGCCCGCTGCTACAGGG 1560  
 QY 1662 CGGTCCCTCAATGATGATACAGCATCTGTGTACACCCATCTGATGTGTTAGGCTTGG 1721  
 DB 1561 CGGTCCCTCAATGATGATACAGCATCTGTGTACACCCATCTGATGTGTTAGGCTTGG 1620  
 QY 1722 CGGTCCCTCAATGATGATACAGCATCTGTGTACACCCATCTGATGTGTTAGGCTTGG 1781  
 DB 1621 CGGTCCCTCAATGATGATACAGCATCTGTGTACACCCATCTGATGTGTTAGGCTTGG 1680  
 QY 1782 CTGGAACCTCACTCGGAGGAGTGCAGAGCTGTGTACAGCTTGTACTATGAGAGGCAAGA 1841  
 DB 1681 CTGGAACCTCACTCGGAGGAGTGCAGAGCTGTGTACAGCTTGTACTATGAGAGGCAAGA 1740  
 QY 1842 AGGCTTCACTGAGCAAGAGGCTGCTCTGTTGAGAGGCTGAGAGGCTGCTGAGCTAT 1901  
 DB 1741 AGGCTTCACTGAGCAAGAGGCTGCTCTGTTGAGAGGCTGAGAGGCTGCTGAGCTAT 1800  
 QY 1902 GAGCTGCTCCGCGCACTGAGCAAGAGTGTGCTGCTAGTGAACGCCGCTTCTGCTGAGGACG 1961  
 DB 1801 GAGCTGCTCCGCGCACTGAGCAAGAGTGTGCTGCTAGTGAACGCCGCTTCTGCTGAGGACG 1860  
 QY 1962 TGGCTAGACAGAGCCCGGAGCAAGAGTGTGCTGCTAGTGAACGCCGCTTCTGCTGAGGACG 2021  
 DB 1861 TGGCTAGACAGAGCCCGGAGCAAGAGTGTGCTGCTAGTGAACGCCGCTTCTGCTGAGGACG 1920  
 QY 2022 AACAGCCGCTACAGCTGACCTTGTGTGGGGCCAGAGAGGCAATCTGAGCTATGCCAAC 2081  
 DB 1921 AACAGCCGCTACAGCTGACCTTGTGTGGGGCCAGAGAGGCAATCTGAGCTATGCCAAC 1980  
 QY 2082 AAGCAGCTGCGGGGCTGTGTGCGCCCACTACTACACCCCTGCTGCGGCTTCTCTGAG 2141  
 DB 1981 AAGCAGCTGCGGGGCTGTGTGCGCCCACTACTACACCCCTGCTGCGGCTTCTCTGAG 2040  
 QY 2142 GCGCTGCTGACAGTGTGCGCCAGGAGCATCCTTCTTCCAAAGCAGCAGTGTGACAAAT 2201  
 DB 2041 GCGCTGCTGACAGTGTGCGCCAGGAGCATCCTTCTTCCAAAGCAGCAGTGTGACAAAT 2100  
 QY 2202 GTCTTCCAACTGAGAGAGGCTTCTGCTGAGCAAGCAGAGTATCCCAAGCCAGCCGCGA 2261  
 DB 2101 GTCTTCCAACTGAGAGAGGCTTCTGCTGAGCAAGCAGAGTATCCCAAGCCAGCCGCGA 2160  
 QY 2262 GAGACACTGTGAGACCTGCGCAAGAGATCTTCTCAATATTACCCCGGCTGGTGGCC 2321  
 DB 2161 GAGACACTGTGAGACCTGCGCAAGAGATCTTCTCAATATTACCCCGGCTGGTGGCC 2220  
 QY 2322 GAGCTTGTGATGATGATGCGACCACTGCGCTTGTCTTCCGCTAATTCAGAGGACAT 2381  
 DB 2221 GAGCTTGTGATGATGATGCGACCACTGCGCTTGTCTTCCGCTAATTCAGAGGACAT 2280  
 QY 2382 TCCAGGGCCCAAGAGCTGAGCAAGATCAAGAGTAAACCAAGGCTGGAGAGAGGCGCCAC 2441  
 DB 2281 TCCAGGGCCCAAGAGCTGAGCAAGATCAAGAGTAAACCAAGGCTGGAGAGAGGCGCCAC 2340  
 QY 2442 GAGCTGCTGTGAGGCTGAGCTGAGGAGGATTTGAGAGAAATGACCTGCTCCACAC 2501  
 DB 2341 GAGCTGCTGTGAGGCTGAGCTGAGGAGGATTTGAGAGAAATGACCTGCTCCACAC 2400  
 QY 2502 ACCCAAGTGTGAGTAAAGTACTGTTCTTCTTCCACTTAAAAAAAAAAAAAAA 2556  
 DB 2401 ACCCAAGTGTGAGTAAAGTACTGTTCTTCTTCCACTTAAAAAAAAAAAAAAA 2455

RESULT 4  
 LOCUS HUMHFS 2446 bp mRNA linear PRI 25-NOV-1996  
 DEFINITION Human alpha-N-acetylglucosaminidase (uHfSD1) mRNA, complete cds.  
 ACCESSION U78464  
 VERSION L78464.1 GI:1479980  
 KEYWORDS  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 2446)  
 AUTHORS Zhao, Z., Yazdani, A., Shen, Y., Sun, Z., Bailey, J., Caskey, C.T. and Lee, C.C.  
 TITLE Molecular dissection of a cosmid from a gene-rich region in 17q21 and characterization of a candidate gene for alpha-N-acetylglucosaminidase with two cDNA isoforms  
 JOURNAL Mamm. Genome 7 (9), 686-690 (1996)  
 MEDLINE 96359157  
 PUBMED 8703123  
 FEATURES  
 source  
 1. .2446  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="adult"  
 1. .2446  
 /gene="uHfSD1"  
 <1. .2446  
 /gene="uHfSD1"  
 <1. .7  
 /gene="uHfSD1"  
 7. .2238  
 /gene="uHfSD1"  
 /codon\_start=1  
 /evidence=experimental  
 /product="alpha-N-acetylglucosaminidase"  
 /protein\_id="AAB36604.1"  
 /db\_xref="GI:1479981"  
 /translation="MEAVAVAAVAVGLILAGAGAGDAREAAVAVLVARLLGPGP  
 AADPSVVERALAAKRLDPTSLGGGAGARVVRGSGTGAALAAHRYLRDCCGCHA  
 WSSQRLRPRPAINVDELTEATPRIRYQVCTDSYGFWMMDMARREIDDMALN  
 GINLALMSGQELIOWRVYLAGLTAEINERFTGFAIAGMGMLHTWGDLPSPW  
 HIKOYLORHVLDMQMSFGMTPLVPAFAHVEAVTRVPPQVAVTKMSGWHFNCYS  
 CSEFLAPEDIPPIIGSLFLRELKEFGDHTYGADEFEMQPPSSPEVYLAATYAV  
 YEAMTAVDEAVVLLGMLFOHOPQPMGAQIRAVLGAVPRGLVLVDLFAESQPYT  
 RTASFGQPPIMWMLNPGNIGLFGALAVNAGVPAARLPNSITWVGMAPEGISGA  
 NEVYSALMEGRDPPVPLAAMWTSFPAARVGVSHPDAGAMRLIPSVNCSSEA  
 CRGHNSPLVRPSLOMTSIVNNSDVEAMRLITSPSLATSPAFRDLIDLTRO  
 AVGLVSLYEEKRSAYLSKELASLIFAGGVLAELPLALDEVLASDSFLLGSWLEQ  
 ARAVAQSEADPFEQNSRYOLTLWGBENILIDYANKQLAGLVANYYTRMRLLFEAL  
 VDSVAQGIPIFOGHPDKNVFQLEQAFVLSKORVPSQPRSDTVDLAKIKLYPGWVA  
 GSW"  
 2239. .2446  
 /gene="uHfSD1"  
 791 g 468 c  
 3' UTR  
 2239. .2446  
 /gene="uHfSD1"  
 791 g 468 c  
 BASE COUNT 412 a 775 c 791 g 468 c  
 ORIGIN  
 Query Match 95.0%; Score 2446; DB 9; Length 2446;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 96 GAGACATGAGAGCGGTGGGCTGGGCGCGGCGGCTGCTTCTCTGAGCGGGGCGC 155  
 DB 1 GAGACATGAGAGCGGTGGGCTGGGCGCGGCGGCTGCTTCTCTGAGCGGGGCGC 60  
 QY 156 GGGGGCGCGGCAAGCGAGCGAGGCGCGGAGAGCGCGGCGGCGGCTGCGGCGCGG 215  
 DB 61 GGGGGCGCGGCAAGCGAGCGAGGCGCGGAGAGCGCGGCGGCGGCTGCGGCGCGG 120  
 QY 216 CTGCTGAGGCGAGGCGCGGCGGCGGCTTCTCGGTGTGAGAGCGCGCTGTGCTGCC 275  
 DB 121 CTGCTGAGGCGAGGCGCGGCGGCGGCTTCTCGGTGTGAGAGCGCGCTGTGCTGCC 180  
 QY 276 AAGCGGGGCTTGGACCTTACAGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 335  
 DB 181 AAGCGGGGCTTGGACCTTACAGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
 QY 336 GAGCTCAAGGCGCTGAGCGGCGGCGGCGGCGGCTGACCGCTTCTGCGGCACTTCTG 395  
 DB 241 GAGCTCAAGGCGCTGAGCGGCGGCGGCGGCGGCTGACCGCTTCTGCGGCACTTCTG 300

QY 396 TGCCAGCTGGCTGGCTCGGCTCTCAGCTGCGCTGCGGCCACTGCCAGCTGCGG 455  
Db 301 TGCACGCTGGCTGGCTCGGCTCTCAGCTGCGCTGCGGCCACTGCCAGCTGCGG 360  
QY 456 GGGAGCTGACCGAGGSCAGCCCAAGGTAACGCTATTACAGAGATGTTGACAGCA 515  
Db 361 GGGAGCTGACCGAGGSCAGCCCAAGGTAACGCTATTACAGAGATGTTGACAGCA 420  
QY 516 AGCTACTCTCTTGGTGGTGGGACTGGGCTGCGGAGGAGATGAGCTGGATGGG 575  
Db 421 AGCTACTCTCTTGGTGGTGGGACTGGGCTGCGGAGGAGATGAGCTGGATGGG 480  
QY 576 CTGATGAGCACTCAACCTGGAGCTGGAGCGGCTGGAGAGGCACTTGGAGCGG 635  
Db 481 CTGATGAGCACTCAACCTGGAGCTGGAGCGGCTGGAGAGGCACTTGGAGCGG 540  
QY 636 TACCTGGCTTGGGCTGACCCAGAGAGATCAATGATCTTTTACTGGCTGCTTC 695  
Db 541 TACCTGGCTTGGGCTGACCCAGAGAGATCAATGATCTTTTACTGGCTGCTTC 600  
QY 696 CTGGCTCTGGGGGCGAATGGGCAACTGGCAACTGGGATGGCCCCCTGCCCCCTTGG 755  
Db 601 CTGGCTCTGGGGGCGAATGGGCAACTGGCAACTGGGATGGCCCCCTGCCCCCTTGG 660  
QY 756 CACATCAAGAGCTTTTACTCTGAGCAACGGGCTGGAGCCAGATGGGCTCTTGGCATG 815  
Db 661 CACATCAAGAGCTTTTACTCTGAGCAACGGGCTGGAGCCAGATGGGCTCTTGGCATG 720  
QY 816 ACCCGAGCTGCTGCTGATTCGCGGGCATGTTCCGAGCTGTCACAGAGTGTTCCT 875  
Db 721 ACCCGAGCTGCTGCTGATTCGCGGGCATGTTCCGAGCTGTCACAGAGTGTTCCT 780  
QY 876 CAGGTCAATGTCAGAAATGGGCAAGTGGGGCACTTTTAACTGTTCCATCTGCTG 935  
Db 781 CAGGTCAATGTCAGAAATGGGCAAGTGGGGCACTTTTAACTGTTCCATCTGCTG 840  
QY 936 TTCTCTTGGCTCGGAGAGACCCCAATTTCCCATCATCGGAGACCTCTTCTGAGAG 995  
Db 841 TTCTCTTGGCTCGGAGAGACCCCAATTTCCCATCATCGGAGACCTCTTCTGAGAG 900  
QY 996 CTGATCAAAAGTGTGGCAAGACACATCTTATGGGGCGACACTTTCAATGATGAC 1055  
Db 901 CTGATCAAAAGTGTGGCAAGACACATCTTATGGGGCGACACTTTCAATGATGAC 960  
QY 1056 CCACTTCTCTCAGAGCCCTCTACCTTGGCGAGCCACCACTGCGTCTATGAGGCGATG 1115  
Db 961 CCACTTCTCTCAGAGCCCTCTACCTTGGCGAGCCACCACTGCGTCTATGAGGCGATG 1020  
QY 1116 ACTGCAATGATCTGAGGCTGTGTGGCTGCTCAAGGCTGCTCTTTCAGACACAGCG 1175  
Db 1021 ACTGCAATGATCTGAGGCTGTGTGGCTGCTCAAGGCTGCTCTTTCAGACACAGCG 1080  
QY 1176 CAGTTCTGGGGGGCGCGCCAGATCAGGGCTGTGTGGAGCTGTGCGCGGCGGCTG 1235  
Db 1081 CAGTTCTGGGGGGCGCGCCAGATCAGGGCTGTGTGGAGCTGTGCGCGGCGGCTG 1140  
QY 1236 CTGATCTGAGCTGTTGTGTCGAGCAAGCTGTGTATACCGCACTGCTCTTTCAG 1295  
Db 1141 CTGATCTGAGCTGTTGTGTCGAGCAAGCTGTGTATACCGCACTGCTCTTTCAG 1200  
QY 1296 GGGCAAGCCCTTCACTGTGTGCAAGCTGTGCAACTTTGGGGGAAACATGCTTTTGA 1355  
Db 1201 GGGCAAGCCCTTCACTGTGTGCAAGCTGTGCAACTTTGGGGGAAACATGCTTTTGA 1260  
QY 1356 GGCCTGAGGCTGTGAACGAGGCGCAAGAGCTGCGGCTCTTTCACCACTGACATG 1415  
Db 1261 GGCCTGAGGCTGTGAACGAGGCGCAAGAGCTGCGGCTCTTTCACCACTGACATG 1320  
QY 1416 GTAGGCAACGGGCACTGGCCCCCGAGGGGATGACCAAGAAAGATGTCTATTCCTCATG 1475  
Db 1321 GTAGGCAACGGGCACTGGCCCCCGAGGGGATGACCAAGAAAGATGTCTATTCCTCATG 1380

QY 1476 GCTGAGCTGGGCTGGCCAAAGAACCAAGTCCCAATTTGGAGCTTGGGTACCACTTT 1535  
Db 1381 GCTGAGCTGGGCTGGCCAAAGAACCAAGTCCCAATTTGGAGCTTGGGTACCACTTT 1440  
QY 1536 GCGGCGCGGCTATGGGGGTCTCCCAACCGGAGCGAGGGGAGGCTGAGGCTACTGCTC 1595  
Db 1441 GCGGCGCGGCTATGGGGGTCTCCCAACCGGAGCGAGGGGAGGCTGAGGCTACTGCTC 1500  
QY 1596 CGAGTGTATCAACTGCTCCGGGGAGGCTGACAGGGGCAACAATCGTAGCCGCTGCTC 1655  
Db 1501 CGAGTGTATCAACTGCTCCGGGGAGGCTGACAGGGGCAACAATCGTAGCCGCTGCTC 1560  
QY 1656 AGGCGGCGGCTCCCAATGATATACAGCAATCTGTATCAACCGATCGATGTGTAG 1715  
Db 1561 AGGCGGCGGCTCCCAATGATATACAGCAATCTGTATCAACCGATCGATGTGTAG 1620  
QY 1716 GCTTGGCGGCTGCTGCTCACTGCTCCCTCCCTGGGCAACAGCCCGCTTCCGCTAC 1775  
Db 1621 GCTTGGCGGCTGCTGCTCACTGCTCCCTCCCTGGGCAACAGCCCGCTTCCGCTAC 1680  
QY 1776 GACTGCTGAGCACTCACTGGCAAGCAAGTCAAGAGCTGTCACTTGTATGAGAG 1835  
Db 1681 GACTGCTGAGCACTCACTGGCAAGCAAGTCAAGAGCTGTCACTTGTATGAGAG 1740  
QY 1836 GCAAGAAAGGCTACTGAGCAAGAGGCTGCTCCCTGTTGAGGGCTGAGGCGTCTG 1895  
Db 1741 GCAAGAAAGGCTACTGAGCAAGAGGCTGCTCCCTGTTGAGGGCTGAGGCGTCTG 1800  
QY 1896 GCTTATGAGCTGCTGCGGCACTGAGCAAGGCTGCTGCTGATGACAGCGCTTCTGCTG 1955  
Db 1801 GCTTATGAGCTGCTGCGGCACTGAGCAAGGCTGCTGCTGATGACAGCGCTTCTGCTG 1860  
QY 1956 GGCAGCTGGCTGAGAGAGGCGCGAGCAAGGCTGAGTGAAGGCGAGGCGGATTTCTAC 2015  
Db 1861 GGCAGCTGGCTGAGAGAGGCGCGAGCAAGGCTGAGTGAAGGCGAGGCGGATTTCTAC 1920  
QY 2016 GAGCAAAACAGCCGCTACCACTGATCTGTGGGGGCGAAGGCAACATCTGACTAT 2075  
Db 1921 GAGCAAAACAGCCGCTACCACTGATCTGTGGGGGCGAAGGCAACATCTGACTAT 1980  
QY 2076 GCGAACAAGAGCTGGGGGCTGTGGGCACTTACACCCCTGCTGGCGGCTTTTC 2135  
Db 1981 GCGAACAAGAGCTGGGGGCTGTGGGCACTTACACCCCTGCTGGCGGCTTTTC 2040  
QY 2136 CTGAGGCGCTGTGTGACAGTGTGGCCAGGGCACTCCCTTCCAAAGCAACAATTTGAC 2195  
Db 2041 CTGAGGCGCTGTGTGACAGTGTGGCCAGGGCACTCCCTTCCAAAGCAACAATTTGAC 2100  
QY 2196 AAAAATGCTTCCAACTGAGAGCAGGCTTGTCTCAGCAAGCAGAGGTACCCAGCGAG 2255  
Db 2101 AAAAATGCTTCCAACTGAGAGCAGGCTTGTCTCAGCAAGCAGAGGTACCCAGCGAG 2160  
QY 2256 CCGGAGAGAGCACTGTGAGCACTGAGCCCAAGAAATCTTCCCAATTTACCCCGGCTG 2315  
Db 2161 CCGGAGAGAGCACTGTGAGCACTGAGCCCAAGAAATCTTCCCAATTTACCCCGGCTG 2220  
QY 2316 GTGGCGGCTTGTGTATGATTTGCACTGAGGCTGTTGTTTCGCTAATTTCAAG 2375  
Db 2221 GTGGCGGCTTGTGTATGATTTGCACTGAGGCTGTTGTTTCGCTAATTTCAAG 2280  
QY 2376 GCAATTCAGAGGCGCGAGAGCTGAGCAAGATCAAGATTAACCAAGGCTGGAGGAGG 2435  
Db 2281 GCAATTCAGAGGCGCGAGAGCTGAGCAAGATTAACCAAGGCTGGAGGAGG 2340  
QY 2436 CCGCAGGCGCTGCTGTGGGGCTGACCTGCGGGGATTTGAGAGGAAATGACTGCGCTCC 2495  
Db 2341 CCGCAGGCGCTGCTGTGGGGCTGACCTGCGGGGATTTGAGAGGAAATGACTGCGCTCC 2400  
QY 2496 ACCAACAACCAATGTGGATTTAAATGATCTGTTTCTTTCCACTT 2541  
Db 2401 ACCAACAACCAATGTGGATTTAAATGATCTGTTTCTTTCCACTT 2446

[illegible]

Db	996	GGAGCGGCACGAGAGGCATCTGGCAGGGGGGTGTAACTGCGCTTGAGGCGTCAACCGAGCAG	1055
QY	664	AGATCAATGAGTTCTTTACTGTCGTCCGCTTCCCTGGCGTGGGGGGCGAATGGGCAACCTGC	723
Db	1056	AGATCAATGAGTTCTTTACTGTCGTCCGCTTCCCTGGCGTGGGGGGCGAATGGGCAACCTGC	1115
QY	724	ACACTGGAGTGGCCCCCTGCCCCCTCTCGGACATCAAGCAGCTTTTACTCTGCAGACC	783
Db	1116	ACACTGGAGTGGCCCCCTGCCCCCTCTCGGACATCAAGCAGCTTTTACTCTGCAGACC	1175
QY	784	GGGTCTCTGGACCAGATGCGCTCCTTGGGCAATGACCCCAATGCTGCTGCAATTGGCGGGGC	843
Db	1176	GGGTCTCTGGACCAGATGCGCTCCTTGGGCAATGACCCCAATGCTGCTGCAATTGGCGGGGC	1235
QY	844	ATGTTCCCGAGGCGTGCACAGGGGTGTTCCTCAAGTCATATGTCACAGAAATGGGCACTT	903
Db	1236	ATGTTCCCGAGGCGTGCATCCAGGGGTGTTCCTCAAGTCATATGTCACAGAAATGGGCACTT	1295
QY	904	GGGGCCACCTTTAACTGTTCTTACTCTGCTCTCTTCTTGCTCTGGCTCGCGAAGACCCCAT	963
Db	1296	GGGGCCACCTTTAACTGTTCTTACTCTGCTCTCTTCTTGCTCTGGCTCGCGAAGACCCCAT	1355
QY	964	TCCCAATCATCGGGAGGCTCTCTCGTGGAGAGCTGATCAAGAAGTTTGGACAGACACA	1023
Db	1356	TCCCAATCATCGGGAGGCTCTCTCGTGGAGAGCTGATCAAGAAGTTTGGACAGACACA	1415
QY	1024	TCTATGGGGCGCACACTTTCAATGAGATGACGCCACCTTCTCAAGACCTCTTACTTG	1083
Db	1416	TCTATGGGGCGCGACACTTTCAATGAGATGACGCCACCTTCTCAAGACCTCTTACTTG	1475
QY	1084	CCGGAAGCACCACTGCGCGCTCTATAGAGGCCATGATGAGATGATCTGAAGGCTGTGGGC	1143
Db	1476	CCGGAAGCACCACTGCGCGCTCTATAGAGGCCATGATGAGATGATCTGAAGGCTGTGGGC	1535
QY	1144	TGCTCCAAAGGTGGCTCTTCCAGACACAGCGCGAGTCTGGGGGCCCGCCAGATCAAGG	1203
Db	1536	TGCTCCAAAGGTGGCTCTTCCAGACACAGCGCGAGTCTGGGGGCCCGCCAGATCAAGG	1595
QY	1204	CTGTGCTGGAGACTGTGTGCCCCGTGTGGCGCTCCTGATCTGTGAACTGTGTTGCTGAAGCC	1263
Db	1596	CTGTGCTGGAGACTGTGTGCCCCGTGTGGCGCTCCTGATCTGTGAACTGTGTTGCTGAAGCC	1655
QY	1264	AGCGTGTATACCCGAGCTGCGCTCTTCCAGAGGCCAGGCCCTTCACTGTGTGATGCTGC	1323
Db	1656	AGCGTGTATACCCGAGCTGCGCTCTTCCAGAGGCCAGGCCCTTCACTGTGTGATGCTGC	1715
QY	1324	ACAACTTTGGGGGAACCAATGCTTTTGGAGCCTTAGAGCTGTGAACGAGAGCCAG	1383
Db	1716	ACAACTTTGGGGGAACCAATGCTTTTGGAGCCTTAGAGCTGTGAACGAGAGCCAG	1775
QY	1384	AAGTGTGCCCCCTTCCCAACTCCACCAATGTATGGGCAAGGGATGAGCCCCCGAGGGCA	1443
Db	1776	AAGTGTGCCCCCTTCCCAACTCCACCAATGTATGGGCAAGGGATGAGCCCCCGAGGGCA	1835
QY	1444	TCAGCCGAAGAGAGTGTATTTCCCTCATGTGCTGAAGCTGGCTGGCGAAAGACCCAG	1503
Db	1836	TCAGCCGAAGAGAGTGTATTTCCCTCATGTGCTGAAGCTGGCTGGCGAAAGACCCAG	1895
QY	1504	TGCCAGATTTGGGCAAGCTTGGGTATCAGACTTTTGGCGCCGGCGGTATGGGGTCTCCAC	1563
Db	1896	TGCCAGATTTGGGCAAGCTTGGGTATCAGACTTTTGGCGCCGGCGGTATGGGGTCTCCAC	1955
QY	1564	CGGACGACAGGGGACGCTGAGAGCTACGTGCTCGGAGTGTGTACAACTGCTCCGGGAGG	1623
Db	1956	CGGACGACAGGGGACGCTGAGAGCTACGTGCTCGGAGTGTGTACAACTGCTCCGGGAGG	2015
QY	1624	CTTGAGGGGGCCAACTGTATGCCCGCTGTCAGGGCGGCGGTCTCTACATGATATACCA	1683
Db	2016	CTTGAGGGGGCCAACTGTATGCCCGCTGTCAGGGCGGCGGTCTCTACATGATATACCA	2075
QY	1684	GCATCTGGTACAACCGATCTGATGTGTTTGAAGCTGAGCGGCTCTCTCATCATCTGCTC	1743
Db	2076	GCATCTGGTACAACCGATCTGATGTGTTTGAAGCTGAGCGGCTCTCTCATCATCTGCTC	2135

QY 1744 CTTCCCTGGGACCAAGCCCGCTTCCGCTACGACTCTGAGACTCTCGGAGGCGAG 1803  
 Db 2136 CTTCCCTGGGACCAAGCCCGCTTCCGCTACGACTCTGAGACTCTCGGAGGCGAG 2195  
 QY 1804 TGCAGAGAGCTGTGAGCTGTGATATATGAGAGGCAAGAGCGCTACCTGAGCAAGAGC 1863  
 Db 2196 TGCAGAGAGCTGTGAGCTGTGATATATGAGAGGCAAGAGCGCTACCTGAGCAAGAGC 2255  
 QY 1864 TGGCTTCCCTGTTGAGGAGCTGAGAGCGCTCTGAGCTATGAGCTCTGAGCGAGCTGAGAG 1923  
 Db 2256 TGGCTTCCCTGTTGAGGAGCTGAGAGCGCTCTGAGCTATGAGCTCTGAGCGAGCTGAGAG 2315  
 QY 1924 AGGAGCTGAGTATGAGAGAGCGCTTCTGAGGAGAGCTGAGTATGAGAGAGCGAGAG 1983  
 Db 2316 AGGAGCTGAGTATGAGAGAGCGCTTCTGAGGAGAGCTGAGTATGAGAGAGCGAGAG 2375  
 QY 1984 CGGAGCTGAGTATGAGAGAGCGCTTCTGAGGAGAGCTGAGTATGAGAGAGCGAGAG 2043  
 Db 2376 CGGAGCTGAGTATGAGAGAGCGCTTCTGAGGAGAGCTGAGTATGAGAGAGCGAGAG 2435  
 QY 2044 TGTGGGGGCGAGAGAGCAATCTGAGTATGAGAGAGAGCGCTGAGGAGGAGTGTGAG 2103  
 Db 2436 TGTGGGGGCGAGAGAGCAATCTGAGTATGAGAGAGAGCGCTGAGGAGGAGTGTGAG 2495  
 QY 2104 CCAACTACTACACCCCTCGCTGAGGAGCTTCTGAGGAGGAGCTGAGTATGAGAGGAG 2163  
 Db 2496 CCAACTACTACACCCCTCGCTGAGGAGCTTCTGAGGAGGAGCTGAGTATGAGAGGAG 2555  
 QY 2164 AGGAGCTGAGTATGAGAGAGCGCTTCTGAGGAGAGCTGAGTATGAGAGAGCGCT 2223  
 Db 2556 AGGAGCTGAGTATGAGAGAGCGCTTCTGAGGAGAGCTGAGTATGAGAGAGCGCT 2615  
 QY 2224 TCGTCTCAGCAAGAGAGGATCCAGAGCGAGCGAGAGAGCACTGTGAGAGCTGAGAG 2283  
 Db 2616 TCGTCTCAGCAAGAGAGGATCCAGAGCGAGCGAGAGAGCACTGTGAGAGCTGAGAG 2675  
 QY 2284 AGAAGATCTTCTCAATATTTAACCAGGAGCTGAGGAGCGAGCTTGTGATGATGAG 2343  
 Db 2676 AGAAGATCTTCTCAATATTTAACCAGGAGCTGAGGAGCGAGCTTGTGATGATGAG 2735  
 QY 2344 ACCAGCTGAGGAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAG 2403  
 Db 2736 ACCAGCTGAGGAGCTTGTGATGATGATGATGATGATGATGATGATGATGAG 2795  
 QY 2404 ACATCAGAGATTAACCAAGGAGCTGAGGAGAGCGAGCGAGCTGAGGAGGAGTGAAG 2463  
 Db 2796 ACATCAGAGATTAACCAAGGAGCTGAGGAGAGCGAGCGAGCTGAGGAGGAGTGAAG 2855  
 QY 2464 TGGGAGGAGATGAGAGGAGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 2523  
 Db 2856 TGGGAGGAGATGAGAGGAGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 2915  
 QY 2524 ACTGTTTCTTCTGAGCT 2541  
 Db 2916 ACTGTTTCTTCTGAGCT 2933

RESULT 6  
 MMU85247 2452 bp mRNA linear ROD 03-DEC-1997  
 LOCUS complete cds.  
 DEFINITION Mus musculus lysosomal alpha-N-acetylglucosaminidase Naglu mRNA,  
 accession U85247  
 VERSION U85247.1 GI:2660687  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2452)  
 REFERENCE  
 AUTHORS Zhao, K.W., Li, H.H. and Neufeld, E.F.  
 TITLE Cloning and expression of mouse gene encoding the lysosomal

JOURNAL  
 Unpublished  
 REFERENCE  
 2 (bases 1 to 2452)  
 AUTHORS  
 Zhao, K.W., Li, H.H. and Neufeld, E.F.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (13-JUN-1997) Biol. Chem., UCLA, 10833 Le Conte Ave, Los  
 Angeles, CA 90095 USA  
 FEATURES  
 source  
 1..2452  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="B6/CDAP1J"  
 /db\_xref="taxon:10090"  
 /tissue\_type="liver"  
 1..2220  
 /EC\_number="3.2.1.50"  
 /function="hydrolyse alpha-N-acetylglucosaminide from  
 glycosaminoglycans"  
 /note="lysosomal alpha-N-acetylglucosaminidase"  
 /codon\_start=1  
 /product="Naglu"  
 /protein\_id="AAB88084.1"  
 /db\_xref="gi:2660688"  
 /translation="MEAGLAVALIGFLLAGSVDEAREKAVRELVRLGPGPA  
 NPLVSVERALADESGDITVSLSGGAGVPLVVRGTVAAAGLRVLRDFGCGVAMS  
 SNOHLMPPLPAVDPDGLTETTPNRYRYQVNCYSYSPVWMDARMEQEDIMNLGI  
 NLALNMGDEALINRVTLALGDTSEIDITFTGAFILAMGRLNLTMDGLPRLSKHL  
 SÖVYLÖRHLIDLRKMSFGMITPLPAFAHVPALITRVPPÖVNLGSGMGNFSCS  
 FLIAPDMPPLIGNFLRELKEFGDHIYGADTFENMOFPFSEPSYLAATTAAYE  
 AMVAVDDAVVLÖGMFLFOHPQFPGSÖIRAVLEAVPRGLVLVDLFLASHPYVMHT  
 ASFGFPIFMCLNFGNGLFGALDVRNRPQALPFSITVVGIGPEGIGONE  
 VVVALMELGMRDPPVDILMAWSSPFIIRRYGSPDPAVAMKLLISVNCSEAS  
 GHNSPLVKRSLQMSNTAVYNSDVEARMLITAPNLTSPAFRYDLDITRÖV  
 OELVSLCYEARKLYLQÖEDLILRAGLLVYKULPTLDELASSHFLGTLWDÖR  
 KAAVSEAPQAFYÖNNSRYÖITLWGEENIÖDYANKÖLAGVADYÖPRMCLFLGTLH  
 SLANGVPPÖHPEKRVFPLEÖAFVYNKKEYPSPRGDITVLSKRIFLKXHPDPSL"  
 <1..377  
 exon  
 378..525  
 exon  
 526..672  
 exon  
 673..758  
 exon  
 759..1015  
 exon  
 1016..>2220  
 BASE COUNT  
 486 a 735 c 709 g 522 t  
 ORIGIN

Query Match 63.1%; Score 1623.6; DB 10; Length 2452;  
 Best Local Similarity 82.8%; Pred. No. 4; le-207;  
 Matches 1864; Conservative 0; Mismatches 380; Indels 6; Gaps 1;

QY 110 GGTGGCGGTGGCCCGCGCGGTGGGGGTCTTCTCTCTGAGCGGAGCGGAGCGGAGCGGAG 169  
 Db 3 GGAAGCGCGGGGCGGTGGCCGTGATCTTGAGGGTTCCTACTCTGAGCTGGGGCTGTGGG 62  
 QY 170 CGACGAGGCGCGGAGCGGCGCGCGCGCGCTGTGGCGCCGCTGCTGGGCGCAG 229  
 Db 63 TGATGAGGCTCGGAGGAGCAAGCGCGGCGGAACTGTGTGTGCTGAGTGGACCGG 122  
 QY 230 CCGCGCGCGGACTTCTCGGTGTGAGTGAAGCGGCTGTGAGTCAAGCGCGGCTTGA 289  
 Db 123 GCGGCGCGGCAATTTCTTGATGTCCGTGAGAGCGGCGCTGAGCAAGTGGGCTGGA 182  
 QY 290 CACTTAAGCTTGGCGGCGGCGGCGGCGGCGGCGGTGGGCTGAGCTTCAAGCGGCGT 349  
 Db 183 CACTTAAGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCTGAGCTTCAAGCGGCGT 242  
 QY 350 GCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 409  
 Db 243 GCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 302  
 QY 410 GTCCGAGCTTCACTGAGTGGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 469  
 Db 303 GTCCGAGCTTCACTGAGTGGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 362

470 GGCCAGCCCAAGTACCGCTATTAACAAGATGTGACACCGCAAGTACTCTTCGT 529  
 363 AACCAACGCCCAAGGATACCGCTATTACAGAAATGTGACACGACAGCTACTCTTCGT 422  
 530 GTGTGGGACTGGGCGCGCTGGAGAGAGATAGACTGATGGCGGTGAATGGCATCAA 589  
 423 GTGTGGGACTGGGCGCGCTGGAGAGATAGACTGATGGCGGTGAATGGCATCAA 482  
 590 CCTGGCACTGGGCGCGCTGGAGAGAGATAGACTGATGGCGGTGAATGGCATCAA 649  
 483 CTGGCTCTGGCTTGAATGGAGAGAGAGATAGACTGATGGCGGTGAATGGCATCAA 542  
 650 CCTGACCCAGAGAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT 709  
 543 CCTGACCCAGAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGAT 602  
 710 AATGGGCACTGACACACTGGAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 769  
 603 CATGGTAACTTGAACACTGGAGATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662  
 770 TTACCTGACAGACCGGATCTGACACAGATGGCTCTCTGAGATAGACCCAGTGTGCC 829  
 663 CTACCTGACAGATGGAATCTGACACCGAGATGGCTCTCTGAGATAGACCCAGTGTGCC 722  
 830 TGCATTCGCGGGGCAATGTTCCTGAGGCTGTCAACAGGCTGTTCCTTGAAGTCAATG 889  
 723 TGCCCTGACAGAGAGATGGCCCAAGGAGATGACACAGGATGTTCACAGATGATG 782  
 890 GAAGATGGGAGTGGGAGCACTTAACTGTCTCTGATCTCTGCTCTCTGCTGCTC 949  
 783 CAATGGGAGATGGGAGCACTTCAACTGCTCTTCAAGTGTCTCTCTGCTGCTC 842  
 950 GGAAGACCCATATTCCTCATATCGAGAGCTCTCTCTGAGAGCTATCAAAAGTT 1009  
 843 AGGAGACCCATATTCCTCATATCGAGAGCTCTCTCTGAGAGCTATCAAAAGTT 902  
 1010 TGAGACAGACCATATTCCTCATATCGAGAGCTCTCTCTGAGAGCTATCAAAAGTT 1069  
 903 TGAGACAGATATATCTATGAGAGCTCTCTCTGAGAGCTATCAAAAGTT 962  
 1070 GGCCTCTCACTTGCAGAGCACTGCTGCTATGAGAGCTATCAAAAGTT 1129  
 963 GGCCTCTCACTTGCAGAGCACTGCTGCTATGAGAGCTATCAAAAGTT 1022  
 1130 TGAAGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1189  
 1023 TGATGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1082  
 1190 CGCCACATCAAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1249  
 1083 CTCTCAATCAAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1142  
 1250 GTTGTGCTGAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1309  
 1143 GTTGTGCTGAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1202  
 1310 CTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1369  
 1203 CTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1262  
 1370 GAAAG 1429  
 1263 GAAAG 1322  
 1430 GGGCCCGGAG 1489  
 1323 AGCCCCCGAG 1382  
 1490 GCGAAG 1549  
 1383 GCGAAG 1442  
 1550 TGGGGTCTCCACCCGAG 1609

1443 CGGGGCTCTCCAGGCTGATTCGCTGAGCTTGGAGAGCTTCTATCAAAATGTGTACA 1502  
 1610 CTGTCTCCGGGAGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1669  
 1503 CTGTCTCCGGGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1562  
 1670 ACAGATGATATACACATCTGTGATCAACAGATGATGATGATGATGATGATGATGAT 1729  
 1563 ACAGATGATATACAGCTGTGATCAACAGATGATGATGATGATGATGATGATGATGAT 1622  
 1730 GCTCAATCTGTCT 1789  
 1623 GTTGAAG 1682  
 1790 CACTGGAG 1849  
 1683 CACCCGCAAG 1742  
 1850 CCTGAGCAAG 1909  
 1743 CCTGAGCAAG 1802  
 1910 GCGGCACTGAG 1969  
 1803 GCTTCACTAGATGAG 1862  
 1970 GCAGAGCCGAG 2029  
 1863 TCAGGAGCCGAG 1922  
 2030 CTACAG 2089  
 1923 CTACAG 1982  
 2090 GCGGAG 2149  
 1983 GCGAG 2042  
 2150 TGACAG 2209  
 2043 TCACAG 2102  
 2210 ACTGAG 2269  
 2103 ACTGAG 2162  
 2270 TGTGAG 2329  
 2163 CGTGAG 2216  
 2330 GTGATGAG 2389  
 2217 GTGAG 2246

RESULT 7  
 HSU43572 10127 bp DNA linear PRI 11-JUN-1996  
 LOCUS HSU43572  
 DEFINITION Human alpha-N-acetylglucosaminidase (NAGLU) gene, complete cds.  
 ACCESSION U43572  
 VERSION U43572.1 GI:1171228  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eubacteria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 10127)  
 Zhao,H.G., Li,H.H., Bach,G., Schmidchen,A. and Neufeld,E.F.  
 The molecular basis of Sanfilippo syndrome type B  
 Proc. Natl. Acad. Sci. U.S.A. 93 (12), 6101-6105 (1996)  
 JOURNAL MEDLINE  
 PUBMED 8650226

REFERENCE 2 (pages 1 to 10127)  
AUTHORS Zhao, H.G., Li, H.H. and Neufeld, E.F.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1995) Hong G. Zhao, Biological Chemistry, UCLA,  
10833 Le Conte Ave., Los Angeles, CA 90095, USA  
FEATURES Location/Qualifiers

/organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /map\_0sosome="17"  
 /chr="17q21"  
 /cell\_type="fibroblast cell"  
 /gene="NAGLU"  
 /gene="NAGLU"  
 join(1085, 1799, 2542, .2689, 3482, .3628, 3813, .3898, 6091, .6347, 8167, .9568)  
 /gene="NAGLU"  
 /note="5' end of mRNA is uncertain"  
 join(1417, .1799, 2542, .2689, 3482, .3628, 3813, .3898, 6091, .6347, 8167, .9377)  
 /EC number="3.2.1.50"  
 /function="one of four enzymes involved in the degradation of heparan sulfate, specifically removes the alpha-N-acetylglucosamine residues"  
 /note="lysosomal enzyme; deficient in Sanfilippo B syndrome"  
 /codon\_start=1  
 /product="alpha-N-acetylglucosaminidase"  
 /protein\_id="AAC50512.1"  
 /db\_xref="GI:1171229"  
 /translation="MEAYVAVAAGVTLILAGAGACADEEREAAYRAIVARILRPGADPQSLRERPLPAVGELETPRKYRYONCTQSYFVWMDMARWERIDWVALGINTALMGSGCAIMQRYVALGLTQAEINEFEGPAFLMGWGNLHTMGDLPSPYHIKQILQHRVLDQKRSFGMTPLPAFAGHPEAVTVPQVNVTHSGHNCNSYVYAMLRADPEPIPIIGSLFELRLKERGTQHVAGDHFNMSPSPSYLAAITTAAYAMLRADPEAVTILQSWLFOHQPGFMGPAIVAVYGRSLVLDLPAISQPVYRTRSGQGFICMLHFNQNGHGLFGALAEVAGSEPAKLFPSITWGTMAPBGISNBYVSLMAELGWRDVPDLAAVNTSFAARVYSHPDGAAMRLLSVNCSGSECGNHSPLVRPRSLQNNTSITWYKSDVFEWMLILTSASLVSFAPFYDLDLTRAAQVSLTYLEEARASAYLSKEIASLIRAGVGLAELLPLADEVLASDSFLLGSMLEAVALAASEADAEFYQNSRYOGLTWGPGNLTIDVANKQLAGLVANYTTRMRLFEALVDSVAGTIPFGQHGDKNVFQLEQAFLVSKRYPSPGRDPTVLAKKIFLKYYPGVVGSW"  
 1800, .2541  
 /gene="NAGLU"  
 /number=1  
 /gene="NAGLU"  
 /note="the sequence for this intron was partly determined by the submitters and partly derived from Genbank  
 Accession Number M84472"  
 /number=2  
 3629, .3812  
 /gene="NAGLU"  
 /number=3  
 3899, .6090  
 /gene="NAGLU"  
 /note="the sequence for this intron was partly determined by the submitters and partly derived from Genbank  
 Accession Number M84472"  
 /number=4  
 6348, .8166  
 /gene="NAGLU"  
 /note="the sequence for this intron was partly determined by the submitters and partly derived from Genbank  
 Accession Number M84472"  
 /number=5  
 2139 a 2846 c 2824 g 2318 t

Query Match	55.4%	Score 1426.6;	DB 9;	Length 10127;
Best Local Similarity	99.7%	Pred. No. 5.2e-181;		
Matches 1429; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

QY	1112	CATGACCTGCACTGATATCTGATAGGCGGTGTGCTGCTCCAGAGGCTGAGCTCTTCCAGACACCA	1171
Db	8156	CTGTCTCAGAGTGAATCTGAAGCTGTGTGTGCTCTCCAAAGCTTGAGCTCTTCCAGCACCA	8215
QY	1172	GCCGCACTTCTGTGGGGGCGCCGCCAGATCAGAGGCTGTGCTGTGGAGCTGTGCCCGTGGCG	1231
Db	8216	GCCGCACTTCTGTGGGGGCGCCGCCAGATCAGAGGCTGTGCTGTGGAGCTGTGCCCGTGGCG	8275
QY	1232	CCCTCCGCTTCTGGAACCTGTGTTGCTGTGAGACCAAGCCTGTGTATATACCCGACCTGCTCCTT	1291
Db	8276	CCCTCCGCTTCTGGAACCTGTGTTGCTGTGAGACCAAGCCTGTGTATATACCCGACCTGCTCCTT	8335
QY	1292	CCAGAGGCGACGCCCTTCATCTGTGTGATGCTGTGCACAACTTTTGGGGGAAACCATGTGCTCTTTT	1351
Db	8336	CCAGAGGCGACGCCCTTCATCTGTGTGATGCTGTGCACAACTTTTGGGGGAAACCATGTGCTCTTTT	8395
QY	1352	TGAGAGCCCTGAGAGCTGTGAAACGGAGGCCAGAAAGCTGGCCGCTCTTCCCACTCCAC	1411
Db	8396	TGAGAGCCCTGAGAGCTGTGAAACGGAGGCCAGAAAGCTGGCCGCTCTTCCCACTCCAC	8455
QY	1412	CATGTATAGGACAGGGACATGAGCCCCCGAGGGGATCAGCCAGAACGAACTGTGTCTATTCCTT	1471
Db	8456	CATGTATAGGACAGGGGATGAGCCCCCGAGGGGATCAGCCAGAACGAACTGTGTCTATTCCTT	8515
QY	1472	CATGAGCTGACCTGGGCTGTGGGAAAGGACCCAGTGGCCAGATTTGGACGCTTGGGTGACGAC	1531
Db	8516	CATGAGCTGAGCTGGGCTGTGGGAAAGGACCCAGTGGCCAGATTTGGACGCTTGGGTGACGAC	8575
QY	1532	CTTTGCCGCCCGGGGTATGAGGCTTCTCCACCCGGAAGCAGAGGGGACAGCTGTGAGGCTACT	1591
Db	8576	CTTTGCCGCCCGGGGTATGAGGCTTCTCCACCCGGAAGCAGAGGGGACAGCTGTGAGGCTACT	8635
QY	1592	GCTCCGGAAGTGTGATACACTGTCTCCGGGAGAGGCTGTGAGGGGCCCAACATGCTACCCGCT	1651
Db	8636	GCTCCGGAAGTGTGATACACTGTCTCCGGGAGAGGCTGTGAGGGGCCCAACATGCTACCCGCT	8695
QY	1652	GGTCAGAGCGGCGCTGCCCTTACAGATGAATACACAGCATCTGTGTCAACCCGATCTGATGTGTT	1711
Db	8696	GGTCAGAGCGGCGCTGCCCTTACAGATGAATACACAGCATCTGTGTCAACCCGATCTGATGTGTT	8755
QY	1712	TGAGGCTGTGCGGTGTGCTCACTGTGCTCCCTCCCTGGCCACCAAGCCCCGACTCCG	1771
Db	8756	TGAGGCTGTGCGGTGTGCTCACTGTGCTCCCTCCCTGGCCACCAAGCCCCGACTCCG	8815
QY	1772	CTACGACCTGTGACCTCACTCGGACAGGACAGTGTGACAGAGCTGTGCTACGTTGTATCTATGA	1831
Db	8816	CTACGACCTGTGTGACCTCACTCGGACAGGACAGTGTGACAGAGCTGTGCTACGTTGTATCTATGA	8875
QY	1832	GGAGGCAAGAAAGCGCTTAACTGTAGAGCAAGAGCTGGGCTCCCTGTGTGAGGGCTGAGAGCGT	1891
Db	8876	GGAGGCAAGAAAGCGCTTAACTGTAGAGCAAGAGCTGGGCTCCCTGTGTGAGGGCTGAGAGCGT	8935
QY	1892	CTTGAGCCTTATAGCTGTGTCGCGGCACTGTGAACGAGGTCTGTGCTATGATGACAGCGCTTCTT	1951
Db	8936	CTTGAGCCTTATAGCTGTGTCGCGGCACTGTGAACGAGGTCTGTGCTATGATGACAGCGCTTCTT	8995
QY	1952	GCTTGGGACGCTGGCTTAAGAGAGGCGCCGAGACAGCGGCACTCAGTGAAGGCCGAGAGCGCATTT	2011
Db	8996	GCTTGGGACGCTGGCTTAAGAGAGGCGCCGAGACAGCGGCACTCAGTGAAGGCCGAGAGCGCATTT	9055
QY	2012	CTACGAGCAGAAACAAGCCGCTAACCACTGACCTTTTGGGGGCGCAGAGGCAACATCTTGA	2071
Db	9056	CTACGAGCAGAAACAAGCCGCTAACCACTGACCTTTTGGGGGCGCAGAGGCAACATCTTGA	9115
QY	2072	CTATGGAACAAGAGAGCTGGCGGGGCTGTGTGGCCAACTATACACCCCTGCTGAGCGGCT	2133
Db	9116	CTATGGAACAAGAGAGCTGGCGGGGCTGTGTGGCCAACTATACACCCCTGCTGAGCGGCT	9175
QY	2132	TTTCTCTGAGAGCGGTGGTTGACACTGTGTGGCCAGAGGCACTCCCTTTCACACGACCAAGTT	2191

Db 9176 TTTCTGAGAGGCGCTGATGACAGTGTGGCCAGGAGCTCCCTTCCAGACGACAGATT 9235  
Qy 2192 TGACAAAATGCTCTCCACTGAGACAGCCCTGTTCTCAGACAGAGGATACCCGAG 2251  
Db 9236 TGACAAAATGCTCTCCACTGAGACAGCCCTGTTCTCAGACAGAGGATACCCGAG 9235  
Qy 2252 CCAGCCGCGAGAGACACTGTGGACCTGCGCCAGAAAGATCTTCTCAATATTAACCCCG 2311  
Db 9236 CCAGCCGCGAGAGACACTGTGGACCTGCGCCAGAAAGATCTTCTCAATATTAACCCCG 9355  
Qy 2312 CTGGGTGGCCGCTCTGTGTGATGATTCGCCACACCTGGGCTCTGTTTCCGCTAATTC 2371  
Db 9356 CTGGGTGGCCGCTCTGTGTGATGATTCGCCACACCTGGGCTCTGTTTCCGCTAATTC 9415  
Qy 2372 CAGGCGAGATTCAGAGGCGCCAGAGCTGACAGACATCAAGAGATTAACCCAGCCTGGAG 2431  
Db 9416 CAGGCGAGATTCAGAGGCGCCAGAGCTGACAGACATCAAGAGATTAACCCAGCCTGGAG 9475  
Qy 2432 CAGGCGCGAGCTGCTGTGTGTGGGTCTGACCTGGGGGGAATTTGAGGGAATGACTGCC 2481  
Db 9476 CAGGCGCGAGCTGCTGTGTGTGGGTCTGACCTGGGGGGAATTTGAGGGAATGACTGCC 9535  
Qy 2492 CTCACACACACACCAAGATGTGGATTAAGTACTGTTTCTTCCACTTAA 2544  
Db 9536 CTCACACACACACCAAGATGTGGATTAAGTACTGTTTCTTCCACTTAA 9588

RESULT 8  
ARI60673 10380 bp DNA linear PAT 17-OCT-2001  
LOCUS AR160673  
DEFINITION Sequence 3 from patent US 6255096.  
ACCESSION AR160673  
VERSION AR160673.1 GI:16225098  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 10380)  
AUTHORS Hopwood, J. Joseph., Scott, H. Steele., Weber, B., Blanch, L. and  
Amson, D. Stewart.  
TITLE Synthetic mammalian .alpha.-n-acetylglucosaminidase and genetic  
sequences encoding same  
JOURNAL Patent: US 6255096-A 3 03-JUL-2001;  
FEATURES Location/Qualifiers  
source 1..10380  
BASE COUNT 2210 a 2952 c 2852 g 2366 t  
ORIGIN

Query Match 55.4%; Score 1426.6; DB 6; Length 10380;  
Best Local Similarity 99.7%; Pred. No. 5.1e-181;  
Matches 1429; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1412 CATGTAGGCAACGCGATGATCCCGAGGCGATCAGCCAGAAAGATGCTATTCCT 1471  
Db 8034 CATGTAGGCAACGCGATGATCCCGAGGCGATCAGCCAGAAAGATGCTATTCCT 8093  
Qy 1472 CATGTAGGCTGGGTGGCGAAAGAACCAATGCTCAATTTTGGAGAGCTGGGTACAG 1531  
Db 8094 CATGTAGGCTGGGTGGCGAAAGAACCAATGCTCAATTTTGGAGAGCTGGGTACAG 8153  
Qy 1532 CTTTGCGCGCCGCGATGAGGCTCTCCACCCGAGCAGAGGCGAGCGTGAAGCTACT 1591  
Db 8154 CTTTGCGCGCCGCGATGAGGCTCTCCACCCGAGCAGAGGCGAGCGTGAAGCTACT 8213  
Qy 1592 GCTCCGAGATGTGTACATCTGCTCGGAGGAGGCTGACAGGCGCCATCTGACCCGCT 1651  
Db 8214 GCTCCGAGATGTGTACATCTGCTCGGAGGAGGCTGACAGGCGCCATCTGACCCGCT 8273  
Qy 1652 GGTCAAGCGCGCGTCCCTACAGATGAATACAGCATCTGTACACCATCTGATGTGT 1711  
Db 8274 GGTCAAGCGCGCGTCCCTACAGATGAATACAGCATCTGTACACCATCTGATGTGT 8333  
Qy 1712 TGAAGCTGAGCGCTGCTGCTCAATCTGCTCTCTCCCTGSCCACCAGCCCGCTTCG 1771  
Db 8334 TGAAGCTGAGCGCTGCTGCTCAATCTGCTCTCTCCCTGSCCACCAGCCCGCTTCG 8393  
Qy 1772 CTACGACCTGCTGAGACTCACTCGGAGGCGAGTGAAGAGCTGTGACGCTGTACTATGA 1831  
Db 8394 CTACGACCTGCTGAGACTCACTCGGAGGCGAGTGAAGAGCTGTGACGCTGTACTATGA 8453  
Qy 1832 GGAGGCAAGAGCGCTACCTTACAGCAAGAGCTGGCTCTCTTGAAGGCTGAGAGCT 1891  
Db 8454 GGAGGCAAGAGCGCTACCTTACAGCAAGAGCTGGCTCTCTTGAAGGCTGAGAGCT 8513  
Qy 1892 CTTGGCTTATGAGCTGCTGCGGCACTGAGCAGAGTGTCTGATGATGAGAGGCTTCTT 1951  
Db 8514 CTTGGCTTATGAGCTGCTGCGGCACTGAGCAGAGTGTCTGATGATGAGAGGCTTCTT 8573  
Qy 1952 GCTGGGAGCTGCTGAGCAGAGGCGGAGCAGGCACTGAGGCTGAGGCGAGGCTT 2011  
Db 8574 GCTGGGAGCTGCTGAGCAGAGGCGGAGCAGGCACTGAGGCTGAGGCGAGGCTT 8633  
Qy 2012 CTACGAGCAGAACAGCGCTACCACTGACCTTGTGGGGGCGAAGAGCACTCTGGA 2071  
Db 8634 CTACGAGCAGAACAGCGCTACCACTGACCTTGTGGGGGCGAAGAGCACTCTGGA 8693  
Qy 2072 CTATGCGCAACAGCAGCTGCGGCGGCTTGGTGGCGCACTACTACACCCCTGCTGGCGCT 2131  
Db 8694 CTATGCGCAACAGCAGCTGCGGCGGCTTGGTGGCGCACTACTACACCCCTGCTGGCGCT 8753  
Qy 2132 TTTCTGAGAGCGCTGATGACAGTGTGGCCAGGCGATCCCTTCCACAGACCACTT 2191  
Db 8754 TTTCTGAGAGCGCTGATGACAGTGTGGCCAGGCGATCCCTTCCACAGACCACTT 8813  
Qy 2192 TGAACAAAATGTTTCCACTGAGAGAGCGCTTCTGCTCAGAGACAGAGTACCCGAG 2251  
Db 8814 TGAACAAAATGTTTCCACTGAGAGAGCGCTTCTGCTCAGAGACAGAGTACCCGAG 8873  
Qy 2252 CCAGCCGCGAGAGACACTGTGACCTGGCCAAAGATCTTCTCAATATTAACCCCG 2311  
Db 8874 CCAGCCGCGAGAGACACTGTGACCTGGCCAAAGATCTTCTCAATATTAACCCCG 8933  
Qy 2312 CTGGGTGGCGGCTCTGTGTGATGATTCGCCACACTGGGCTTGTCTCCGCTAATTC 2371  
Db 8934 CTGGGTGGCGGCTCTGTGTGATGATTCGCCACACTGGGCTTGTCTCCGCTAATTC 8993  
Qy 2372 CAGGCGAGATTCAGAGGCGCCAGAGCTGACAGACATCAAGAGATTAACCCAGCCTGGAG 2431  
Db 8994 CAGGCGAGATTCAGAGGCGCCAGAGCTGACAGACATCAAGAGATTAACCCAGCCTGGAG 9053  
Qy 2432 GAGGCGCCAGCGCTGCTGTGTGGGTCTGACCTGGGGGGAATTTGAGAGAAATGACTGCC 2491  
Db 9054 GAGGCGCCAGCGCTGCTGTGTGGGTCTGACCTGGGGGGAATTTGAGAGAAATGACTGCC 9113

QY 2492 CTCACACACACACCAAGTGGATTAAAGTACTGTTTCTTCCACTTAA 2544  
 Db 9114 CTCACACACACACCAAGTGGATTAAAGTACTGTTTCTTCCACTTAA 9166

RESULT 9  
 HSU34879 46610 bp DNA linear PRI 14-JAN-1999  
 LOCUS Human 17-beta-hydroxysteroid dehydrogenase (BDH17B2) gene, complete  
 DEFINITION cds.  
 ACCESSION U34879  
 VERSION U34879.1 GI:975694  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 46610)  
 Zhao,Z., Yazdani,A., Shen,Y., Sun,Z., Bailey,J., Caskey,C.T. and  
 Lee,C.C.  
 REFERENCE Molecular dissection of a cosmid from a gene-rich region in 17q21  
 and characterization of a candidate gene for  
 alpha-N-acetylglucosaminidase with two cDNA isoforms  
 Mamm. Genome 7 (9), 666-690 (1996)  
 8703123  
 MEDLINE  
 PUBMED  
 2 (bases 1 to 46610)  
 Shen,Y. and Gibbs,R.A.  
 REFERENCE Direct Submission  
 TITLE Submitted (25-AUG-1995) Ying Shen and Richard A. Gibbs, Molecular &  
 Human Genetics, Baylor College of Medicine, One Baylor Plaza,  
 Houston, TX 77030, USA

FEATURES  
 source  
 1. .46610  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17q21.1"  
 /clone="140C10"  
 /clone\_1b="Los Alamos Cosmid Library: LANCO17"  
 /complement(356. .528)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(529. .806)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(1082. .1412)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(1414. .1749)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 2374. .2717  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 complement(2868. .3078)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 3232. .3631  
 /rpt\_family="MER9"  
 /rpt\_type=direct  
 complement(3633. .3882)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 4239. .4535  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 4668. .4922  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 complement(5694. .5924)  
 /rpt\_family="Alu"

repeat\_unit  
 /rpt\_type=inverted  
 complement(5958. .6237)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(7197. .7485)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(7489. .7796)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(7804. .8085)  
 /rpt\_family="MSAT1"  
 /rpt\_type=inverted  
 8900. .9066  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 complement(9119. .9309)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(9310. .9597)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(9599. .9742)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 12373. .12557  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 12635. .34401  
 /note="region overlaps with GenBank Accession Number  
 M84472"  
 14510. .14805  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 14839. .15086  
 /rpt\_family="L1"  
 15094. .15389  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 complement(16583. .16749)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(16750. .17039)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(17179. .17466)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(19249. .19418)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(19419. .19713)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 19918. .20162  
 /rpt\_family="MER21"  
 /rpt\_type=direct  
 complement(20167. .20504)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 20748. .20899  
 /rpt\_family="MER40"  
 /rpt\_type=direct  
 23522. .23755  
 /rpt\_family="Alu"  
 /rpt\_type=direct  
 complement(23988. .24293)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 complement(25635. .35976)  
 /rpt\_family="Alu"  
 /rpt\_type=inverted  
 25854. .26145

		/rpt_family="Alu"
		/rpt_type=direct
		26166..26454
		/rpt_family="Alu"
		/rpt_type=direct
		complement(26507..26693)
		/rpt_family="Alu"
		/rpt_type=inverted
		26645..29892
		/gene="EDH17B2"
mRNA		join(26645..27709,27803..27970,28118..28297,28473..28566,
		29084..29261,29348..29892)
		/gene="EDH17B2"
		/product="17-beta-hydroxysteroid dehydrogenase"
		26645..27709
		/gene="EDH17B2"
		/number=1
		26900..27070
		/gene="EDH17B2"
		/rpt_family="MER40"
		/rpt_type=direct
		join(27613..27709,27803..27970,28118..28297,28473..28566,
CDS		29084..29261,29348..29892)
		/gene="EDH17B2"
		/codon_start=1
		/product="17-Beta-Hydroxysteroid Dehydrogenase"
		/protein_id="AAD05013.1"
		/db_xref="gi:975895"
		/translation="MARTVVLTGSSSGIGHLAVRLASDPSPQSFKYATIRDLKTOG
		RUMEARACPPGSLETTLQDVRDKSVAAAREVTGRVDVLVCNAGLLPAPLEA
		LGEBAVASLDVNNAVGTITRMQLQAFIPMKRRGGSRVLTGVGLMGLPFNDIVCAASK
		FALSGLCESLAVLLIPEGVHLSTIECGVHTAEWEKVIGSBEVLDRDTINTFRFQG
		VTAHRREYEGDVDPKAEEAGAAGAGAGADEAGRASVADPELGDDPAARQ"
exon		27803..27970
		/gene="EDH17B2"
		/number=2
		28118..28297
exon		/gene="EDH17B2"
		/number=3
		28473..28566
exon		/gene="EDH17B2"
		/number=4
		29084..29261
exon		/gene="EDH17B2"
		/number=5
		29348..29892
exon		/gene="EDH17B2"
		/number=6
		29751..30002
repeat_unit		/rpt_family="Alu"
		/rpt_type=direct
repeat_unit		complement(30221..30520)
		/rpt_family="Alu"
repeat_unit		/rpt_type=inverted
		32086..32377
		/rpt_family="Alu"
		/rpt_type=direct
		32380..32705
		/rpt_family="Alu"
		/rpt_type=direct
		Matches 1429; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Query Match	55.4%; Score 1426.6; DB 9; Length 46610;	
Best Local Similarity	99.7%; Pred. No.3.6e-18.1;	
MATCHES 1429; Conserved	0; Mismatches 4; Indels 0; Gaps 0;	
DB	1112 CATGACTCAGTGAATACTAGAGCTGTGTGCTGCCAGGCTGCCTTTCCAGCACCA	1171
DB	17694 COTGTCACAAGTGATGATTGAGGCTGTGTGCTGCTCCAAGGCTGCCTTTCCAGCACCA	17753
DY	1172 GCCGAGTTTCGGGGCCCCGCCCAATCAGGCTGTGCTGGGAGCTGTGCCCCGTGGCCG	1231
DY	17754 GCCGAGTTTCGGGGCCCCGCCCATCAGGCTGTGCTGGGAGCTGTGCCCCGTGGCCG	17813

[illegible]

QY 2312 CTGGGTCGGCGGCTCTTGGTGAATGATTCGCCACCACTGGGCTTGTTCCTCCGTAATTC 23171  
 DB 18894 CTGGGTGGCGCGCTCTTGGTGAATGATTCGCCACCACTGGGCTTGTTCCTCCGTAATTC 18953  
 QY 2372 CAGGGCAGATTCGAGGGGCGCCAGAGCTGAGACATACAGAGGATTAACCCAGGCGCTGGGAG 2431  
 DB 18954 CAGGGCAGATTCGAGGGGCGCCAGAGCTGAGACATACAGAGGATTAACCCAGGCGCTGGGAG 19013  
 QY 2432 GAGGCCCCCAGCGCTGTGTGTGGGCTGACCTGGGCGGATTTGGAGGGAATGAACCTGCC 2491  
 DB 19014 GAGGCCCCCAGCGCTGTGTGTGGGCTGACCTGGGCGGATTTGGAGGGAATGAACCTGCC 19073  
 QY 2492 CTCACCCACCCACCAAGATGTGGATTAAGTACTGTTTCTTCCACTTAA 2544  
 DB 19074 CTCACCCACCCACCAAGATGTGGATTAAGTACTGTTTCTTCCACTTAA 19126  
 RESULT 10  
 AC067852/c 179146 bp DNA linear PRI 27-FEB-2003  
 LOCUS AC067852 Homo sapiens chromosome 17, clone RP11-400F19, complete sequence.  
 DEFINITION AC067852 GI:28570403  
 VERSION AC067852.23  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,  
 Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,D.S.,  
 Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Coyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,  
 Howland,J., Iliev,I., Johnson,R., Jones,C., Kann,L., Kartas,A.,  
 Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lenczky,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meidrum,J., Meneus,L., Mihova,T., Mirman,C., Mlenga,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-APR-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE  
 AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouellell,A., Allen,N.,  
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
 Boguslavsky,L., Bouckgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
 Collymore,A., Cook,A., Cooke,P., Corum,B., Dearrellano,K.,  
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Fero,S.,  
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,  
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
 Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,  
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
 Meidrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,  
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,  
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
 Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,  
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-JAN-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE  
 AUTHORS

Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,  
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
 Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,  
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-JAN-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 REFERENCE  
 AUTHORS

COMMENT  
 JOURNAL  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS

All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: M18R  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submission@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L9335  
 Center clone name: 400\_F\_19  
 -----  
 FEATURES  
 source  
 Location/Qualifiers  
 1..179146  
 /organism="Homo sapiens"  
 /mol\_type="Genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17"  
 /clone="RP11-400F19"  
 /clone\_idb="RP11-400F19 Human Male BAC"  
 138..440  
 /rpt\_family="AluJb"  
 962..1001  
 /rpt\_family="GC\_rich"  
 1058..1081  
 /rpt\_family="GC\_rich"  
 1327..1382  
 /rpt\_family="CT-rich"  
 2432..2469  
 /rpt\_family="GC\_rich"  
 2902..2934  
 /rpt\_family="AGGGGAGn"  
 complement(3214..3365)  
 /rpt\_family="MR3"  
 complement(5216..5317)  
 /rpt\_family="MR"  
 5700..5721  
 /rpt\_family="GC\_rich"  
 complement(6114..6215)



QY 2132 TTCTCTGAGAGCGCTGTTGACAGTGTGSCCCAGGCGATCCCTTCCACAGACCAAGTT 2191  
 DB 133991 TTCTCTGAGAGCGCTGTTGACAGTGTGSCCCAGGCGATCCCTTCCACAGACCAAGTT 133992  
 QY 2192 TGACAAAATGTCTTCCACTGAGAGCGCTTCTTCTGACAGACAGAGTATCCCGAG 2251  
 DB 133931 TGACAAAATGTCTTCCACTGAGAGCGCTTCTTCTGACAGACAGAGTATCCCGAG 133872  
 QY 2252 CCAGCCCGAGAGAGACACTGTGACCTGAGCCAGAGATCTTCTTCCAAAATATATACCCCG 2311  
 DB 133871 CCAGCCCGAGAGAGACACTGTGACCTGAGCCAGAGATCTTCTTCCAAAATATATACCCCG 133812  
 QY 2312 CTGAGTGGCGCGCTCTTGTGATGATTCGCCACCACTGAGCCTTGTCTTCCGCTATTC 2371  
 DB 133811 CTGAGTGGCGCGCTCTTGTGATGATTCGCCACCACTGAGCCTTGTCTTCCGCTATTC 133752  
 QY 2372 CAGGCGAGATTCAGAGGCGCCAGAGCTGAGACAGACATCAGAGATACCAAGCCTGAGGAG 2431  
 DB 133751 CAGGCGAGATTCAGAGGCGCCAGAGCTGAGACAGACATCAGAGATACCAAGCCTGAGGAG 133692  
 QY 2432 GAGGCGCCAGAGCTGCTGTGATGAGTCTGACCTGAGGAGGAGATTCAGAGGAGATGACCTGCC 2491  
 DB 133691 GAGGCGCCAGAGCTGCTGTGATGAGTCTGACCTGAGGAGGAGATTCAGAGGAGATGACCTGCC 133632  
 QY 2492 CTCCACACACCAAGAGTGTGGATTAAGTACTGTTTCTTCCACTTAA 2544  
 DB 133631 CTCCACACACCAAGAGTGTGGATTAAGTACTGTTTCTTCCACTTAA 133579

RESULT 11  
 ACC002537 114531 bp DNA linear HTG 18-APR-1998  
 LOCUS ACC002537/c Homo sapiens chromosome 17 clone hRPC64M4 map 17, \*\*\* SEQUENCING  
 DEFINITION IN PROGRAMS \*\*\* 6 unordered pieces.  
 ACC002537  
 ACC002537.1 GI:3063511  
 HTG: HTG PHASE1.  
 ORGANISM Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 114531)  
 Birren, B., Fasman, K., McKernan, K., Nusbaum, C., Richardson, P. and  
 Lander, E.  
 Homo sapiens chromosome 17, clone hRPC64M4  
 Unpublished  
 2 (bases 1 to 114531)  
 Hawkins, J.L., Birren, B.W., Fasman, K.H., Nusbaum, C., Lander, E.S.,  
 McKernan, K., Munro, C., Richardson, P., Barna, N., Chang, A., Cooke, P.,  
 Daly, M.J., Devon, K., Dewar, K., Forrest, C., Gage, D., Gerbarg, K.,  
 Hagos, B., Harris, K., Huang, J., Hui, L., Jacotot, L., Kirby, A.,  
 Lane, M., Mackenzie, J., Margulis, N., McDermott, J., Molla, M.,  
 Morrow, J., Nachman, A., Naylor, J., Nusbaum, C., O'Connor, T.,  
 Olotu, A., Peterson, K., Reeve, M.P., Roberts, D., Rollins, G.,  
 Stillewell, J., Stone, C., Strickland, C., Sydney, K., Tang, J.,  
 Wilmer, F., Zemtseva, I. and Zody, M.  
 Direct Submission  
 Submitted (12-SEP-1997) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 18, 1998 this sequence version replaced GI:2992495.  
 All repeats were identified using RepeatMasker; Smt, A.F.A. &  
 Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 65733: contig of 65733 bp in length  
 \* gap of unknown length

\* 65734 76180: contig of 10447 bp in length  
 \* 76181 95382: contig of 19202 bp in length  
 \* 95383 98662: contig of 3280 bp in length  
 \* 98663 106933: contig of 8271 bp in length  
 \* 106934 114531: contig of 7598 bp in length.  
 \* Location/Qualifiers  
 1..114531  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17"  
 /clone="hRPC64M4"  
 /clone\_1db="Peter de Jong/ human PAC library"  
 BASE COUNT 31283 a 27342 c 26578 g 28985 t 343 others  
 ORIGIN  
 Query Match 55.0%; Score 1415.6; DB 2; Length 114531;  
 Best Local Similarity 99.7%; Pred. No. 8.7e-180;  
 Matches 1429; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
 1112 CATGACTGCAATGATATG-AGGCTGTGTGCTGCTTCCAAAGCTGTGCTTCCAGACACC 1170  
 DB 4953 CCGTTCACAGATGATATGATGAGCTGTGTGCTTCCAAAGCTGTGCTTCCAGACACC 4894  
 QY 1171 AGCCGCAATTCAGGAGGCGCCCGCCAGATCAGAGGCTGTGTGCTGAGAGCTGTGCGCCG 1230  
 DB 4893 AGCCGCAATTCAGGAGGCGCCCGCCAGATCAGAGGCTGTGTGCTGAGAGCTGTGCGCCG 4834  
 QY 1231 GCTTCTGTGTGAGCTGTTGTGTGAGAGCCAGCTGTGTATACCGGCACTGCTTCT 1290  
 DB 4833 GCTTCTGTGTGAGCTGTTGTGTGAGAGCCAGCTGTGTATACCGGCACTGCTTCT 4774  
 QY 1291 TCCAGGCGCAGCCCTTCACTGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1350  
 DB 4773 TCCAGGCGCAGCCCTTCACTGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 4714  
 QY 1351 TTGAGACCTTGAAGAGCTGTGAGAGGAGCCAGAGCTGCCGCTTCTTCCCACTCCA 1410  
 DB 4713 TTGAGACCTTGAAGAGCTGTGAGAGGAGCCAGAGCTGCCGCTTCTTCCCACTCCA 4654  
 QY 1411 CCATGTATGACAGCGGATGTGCGCCCGGAGGATCAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1470  
 DB 4653 CCATGTATGACAGCGGATGTGCGCCCGGAGGATCAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 4594  
 QY 1471 TCATGTGCTGAGCTGTGAGAGGAG 1530  
 DB 4593 TCATGTGCTGAGCTGTGAGAGGAG 4534  
 QY 1531 GCTTGTGCGCCCGGCGGTATGAGGCTTCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1590  
 DB 4533 GCTTGTGCGCCCGGCGGTATGAGGCTTCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4474  
 QY 1591 TGCTCCGAGTGTGATCACTGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1650  
 DB 4473 TGCTCCGAGTGTGATCACTGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4414  
 QY 1651 TGGTCAAGCGGCGCTGCTTCAAGATGAATACAGAGATTTGTATACCAAGATCTGATGTGT 1710  
 DB 4413 TGGTCAAGCGGCGCTGCTTCAAGATGAATACAGAGATTTGTATACCAAGATCTGATGTGT 4354  
 QY 1711 TTGAGAGCTGAGGAGGTGTGCTGATCATGTGCTTCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 1770  
 DB 4353 TTGAGAGCTGAGGAGGTGTGCTGATCATGTGCTTCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 4294  
 QY 1771 GCTACAGCTTCTGAGACCTCACTCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1830  
 DB 4293 GCTACAGCTTCTGAGACCTCACTCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4234

```

OY 1831 AGAGGCGAAGAGCGCTTACCTGAGCAGAGCTGCGCTTCCCTGTTGAGGCGTGAAGCG 1890
Db 4233 AGGAGCGAAGAGCGCTTACCTGAGCAGAGCTGCGCTTCCCTGTTGAGGCGTGAAGCG 4174
OY 1891 TCCCTGGCTATGAGCTGTGCGCGCAGCTGAGCAGAGCTGCGCTTCCCTGTTGAGGCGTGAAGCG 1950
Db 4173 TCCCTGGCTATGAGCTGTGCGCGCAGCTGAGCAGAGCTGCGCTTCCCTGTTGAGGCGTGAAGCG 4114
OY 1951 TCGTGGGAGCTGTGCTAGAGCAGGCGCCGAGCAGCGGAGTCACTAGTGAAGCGCGGCGATT 2010
Db 4113 TCGTGGGAGCTGTGCTAGAGCAGGCGCCGAGCAGCGGAGTCACTAGTGAAGCGCGGCGATT 4054
OY 2011 TCTAGAGCAGAGCAGCGCGCTTACCACTGACCTTGTGGGCGCCAGAAAGGCAATCTCTGG 2070
Db 4053 TCTAGAGCAGAGCAGCGCGCTTACCACTGACCTTGTGGGCGCCAGAAAGGCAATCTCTGG 3994
OY 2071 ACTATGCGCAGAGCAGCTGCGGCGTGTGGTGGCACTACTACACCCCTGCTGCGCGC 2130
Db 3993 ACTATGCGCAGAGCAGCTGCGGCGTGTGGTGGCACTACTACACCCCTGCTGCGCGC 3934
OY 2131 TTTTCTTGGAGCGCTGTGTTACAGTGTGGCCCAAGGCGATCCCTTCCAAAGCAGCAGT 2190
Db 3933 TTTTCTTGGAGCGCTGTGTTACAGTGTGGCCCAAGGCGATCCCTTCCAAAGCAGCAGT 3874
OY 2191 TTGACAAATATGTCTTCACTGAGCAGCGCTTGTTCAGCAGCAGAGGATCCCA 2250
Db 3873 TTGACAAATATGTCTTCACTGAGCAGCGCTTGTTCAGCAGCAGAGGATCCCA 3814
OY 2251 GCCAGCGCGAGAGAGACACTGTGGAAGCTGCGCAAGAAAGTCTTCTCAATATTAACCCG 2310
Db 3813 GCCAGCGCGAGAGAGACACTGTGGAAGCTGCGCAAGAAAGTCTTCTCAATATTAACCCG 3754
OY 2311 GCTGGGTTGCGCGCTTGTGATGATGATGTCGCCACCACTGGGCGCTTTCCTGCTAAT 2370
Db 3753 GCTGGGTTGCGCGCTTGTGATGATGATGTCGCCACCACTGGGCGCTTTCCTGCTAAT 3694
OY 2371 CCAAGGCGAGTTCAGAGGCGCCAGAGCTGAGCAGACATCAGATTAACCAAGGCGCTGGA 2430
Db 3693 CCAAGGCGAGTTCAGAGGCGCCAGAGCTGAGCAGACATCAGATTAACCAAGGCGCTGGA 3634
OY 2431 GGAAGCGCGAGGCTGTGTTGAGTGTGCTGCTGCGGCGGATTTGAGAGGAAATGACTGC 2490
Db 3633 GGAAGCGCGAGGCTGTGTTGAGTGTGCTGCTGCGGCGGATTTGAGAGGAAATGACTGC 3574
OY 2491 CCTCCACCAACCAACCAAGTGTGAGTAACTGTTTCTTCCACTTAA 2544
Db 3573 CCTCCACCAACCAACCAAGTGTGAGTAACTGTTTCTTCCACTTAA 3520

RESULT 12
HUM17BHYD 21764 bp DNA linear PRI 06-DEC-1899
LOCUS Human 17-beta-hydroxysteroid dehydrogenase (EDH1B1 and EDH17B2)
DEFINITION
ACCESSION M84472.1 GI:806392
VERSION M84472.1
KEYWORDS 17-beta-hydroxysteroid dehydrogenase; Alu repeat; estradiol
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Peltokeho,H., Isomaa,V., Maenlausta,O. and Vilho,R.
TITLE Complete amino acid sequence of human placental 17
JOURNAL beta-hydroxysteroid dehydrogenase deduced from cDNA
MEDLINE FBS Lett. 239 (1), 73-77 (1988)
PUBMED 89031223
REFERENCE 2 (sites)
AUTHORS The,V., Labrie,C., Zhao,H.F., Couet,J., Lachance,Y., Simard,J.,
TITLE Leblanc,G., Cote,J., Berte,D., Gagne,R. and Labrie,F.
Characterization of cDNAs for human estradiol 17 beta-dehydrogenase

```

```

JOURNAL and assignment of the gene to chromosome 17: evidence of two mRNA
MEDLINE species with distinct 5'-termini in human placenta
PUBMED Mol. Endocrinol. 3 (8), 1301-1309 (1989)
REFERENCE 89384667
AUTHORS 279584
TITLE 3 (sites)
JOURNAL Luu-The,V., Labrie,C., Simard,J., Lachance,Y., Zhao,H.F., Couet,J.,
MEDLINE Leblanc,G. and Labrie,F.
PUBMED Structure of two in tandem human 17 beta-hydroxysteroid
TITLE dehydrogenase genes
JOURNAL Mol. Endocrinol. 4 (2), 268-275 (1990)
MEDLINE 90231340
REFERENCE 233005
AUTHORS 4 (bases 1 to 21764)
TITLE Peltokeho,H., Isomaa,V. and Vilho,R.
JOURNAL Genomic organization and DNA sequences of human 17
MEDLINE beta-hydroxysteroid dehydrogenase genes and flanking regions.
PUBMED Localization of multiple Alu sequences and putative cis-acting
AUTHORS elements
JOURNAL Eur. J. Biochem. 209 (1), 459-466 (1992)
MEDLINE 93011163
PUBMED 1327779
COMMENT On May 11, 1995 this sequence version replaced gi:177126.
[4] reports bases 6501-21788.
FEATURES
source Location/Qualifiers
1. 21764
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17q11-q21"
/cell_type="leukocyte"
883..21598
/rpt_family="Alu"
/rpt_type=dispersed
883..1163
/rpt_family="Alu"
/rpt_type=direct
1874..2154
/rpt_family="Alu"
/rpt_type=direct
2457..2736
/rpt_family="Alu"
/rpt_type=direct
complement(3960..4104)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(4122..4404)
/rpt_family="Alu"
/rpt_type=direct
complement(4405..4538)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(4539..4834)
/rpt_family="Alu"
/rpt_type=direct
complement(4835..5130)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(5131..5426)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(5427..5722)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(5723..6018)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(6019..6314)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(6315..6610)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(6611..6906)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(6907..7202)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(7203..7500)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(7501..7796)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(7797..8092)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(8093..8388)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(8389..8684)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(8685..8980)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(8981..9276)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(9277..9572)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(9573..9868)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(9869..10164)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(10165..10460)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(10461..10756)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(10757..11052)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(11053..11348)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(11349..11644)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(11645..11940)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(11941..12236)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(12237..12532)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(12533..12828)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(12829..13124)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(13125..13420)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(13421..13716)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(13717..14012)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(14013..14308)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(14309..14604)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(14605..14900)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(14901..15196)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(15197..15492)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(15493..15788)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(15789..16084)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(16085..16380)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(16381..16676)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(16677..16972)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(16973..17268)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(17269..17564)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(17565..17860)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(17861..18156)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(18157..18452)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(18453..18748)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(18749..19044)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(19045..19340)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(19341..19636)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(19637..19932)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(19933..20228)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(20229..20524)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(20525..20820)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(20821..21116)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(21117..21412)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(21413..21708)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(21709..22004)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(22005..22300)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(22301..22596)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(22597..22892)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(22893..23188)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(23189..23484)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(23485..23780)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(23781..24076)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(24077..24372)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(24373..24668)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(24669..24964)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(24965..25260)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(25261..25556)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(25557..25852)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(25853..26148)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(26149..26444)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(26445..26740)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(26741..27036)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(27037..27332)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(27333..27628)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(27629..27924)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(27925..28220)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(28221..28516)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(28517..28812)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(28813..29108)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(29109..29404)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(29405..29700)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(29701..30000)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(30001..30296)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(30297..30592)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(30593..30888)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(30889..31184)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(31185..31480)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(31481..31776)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(31777..32072)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(32073..32368)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(32369..32664)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(32665..32960)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(32961..33256)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(33257..33552)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(33553..33848)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(33849..34144)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(34145..34440)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(34441..34736)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(34737..35032)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(35033..35328)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(35329..35624)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(35625..35920)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(35921..36216)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(36217..36512)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(36513..36808)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(36809..37104)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(37105..37400)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(37401..37696)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(37697..37992)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(37993..38288)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(38289..38584)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(38585..38880)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(38881..39176)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(39177..39472)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(39473..39768)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(39769..40064)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(40065..40360)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(40361..40656)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(40657..40952)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(40953..41248)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(41249..41544)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(41545..41840)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(41841..42136)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(42137..42432)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(42433..42728)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(42729..43024)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(43025..43320)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(43321..43616)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(43617..43912)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(43913..44208)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(44209..44504)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(44505..44800)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(44801..45096)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(45097..45392)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(45393..45688)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(45689..45984)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(45985..46280)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(46281..46576)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(46577..46872)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(46873..47168)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(47169..47464)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(47465..47760)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(47761..48056)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(48057..48352)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(48353..48648)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(48649..48944)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(48945..49240)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(49241..49536)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(49537..49832)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(49833..50128)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(50129..50424)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(50425..50720)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(50721..51016)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(51017..51312)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(51313..51608)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(51609..51904)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(51905..52200)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(52201..52496)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(52497..52792)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(52793..53088)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(53089..53384)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(53385..53680)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(53681..53976)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(53977..54272)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(54273..54568)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(54569..54864)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(54865..55160)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(55161..55456)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(55457..55752)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(55753..56048)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(56049..56344)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(56345..56640)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(56641..56936)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(56937..57232)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(57233..57528)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(57529..57824)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(57825..58120)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(58121..58416)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(58417..58712)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(58713..59008)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(59009..59304)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(59305..59600)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(59601..59896)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(59897..60192)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(60193..60488)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(60489..60784)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(60785..61080)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(61081..61376)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(61377..61672)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(61673..61968)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(61969..62264)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(62265..62560)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(62561..62856)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(62857..63152)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(63153..63448)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(63449..63744)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(63745..64040)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(64041..64336)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(64337..64632)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(64633..64928)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(64929..65224)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(65225..65520)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(65521..65816)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(65817..66112)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(66113..66408)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(66409..66704)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(66705..67000)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(67001..67296)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(67297..67592)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(67593..67888)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(67889..68184)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(68185..68480)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(68481..68776)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(68777..69072)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(69073..69368)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(69369..69664)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(69665..69960)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(69961..70256)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(70257..70552)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(70553..70848)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(70849..71144)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(71145..71440)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(71441..71736)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(71737..72032)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(72033..72328)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(72329..72624)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(72625..72920)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(72921..73216)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(73217..73512)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(73513..73808)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(73809..74104)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(74105..74400)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(74401..74696)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(74697..74992)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(74993..75288)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(75289..75584)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(75585..75880)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(75881..76176)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(76177..76472)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(76473..76768)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(76769..77064)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(77065..77360)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(77361..77656)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(77657..77952)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(77953..78248)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(78249..78544)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(78545..78840)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(78841..79136)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(79137..79432)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(79433..79728)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(79729..80024)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(80025..80320)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(80321..80616)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(80617..80912)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(80913..81208)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(81209..81504)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(81505..81800)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(81801..82096)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(82097..82392)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(82393..82688)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(82689..82984)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(82985..83280)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(83281..83576)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(83577..83872)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(83873..84168)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(84169..84464)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(84465..84760)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(84761..85056)
/note="right arm"
/rpt_family="Alu"
/rpt_type=direct
complement(85057..85352)
/note="left arm"
/rpt_family="Alu"
/rpt_type=direct
complement(
```

gene 7730..13913  
/gene="EDH17B1"  
repeat\_unit 7730..13913  
/gene="EDH17B1"  
misc\_difference 7802..7805  
/note="G00-119-860"  
/gene="EDH17B1"  
misc\_difference 7809..7815  
/note="present in clone lambda17-HSD5; absent in clone lambda17-HSD3"  
/gene="EDH17B1"  
prim\_transcript 7816..11014  
/note="G00-119-860"  
/gene="EDH17B1"  
misc\_difference 8095  
/gene="EDH17B1"  
/note="G in clone lambda17-HSD5; A in clone lambda17-HSD3;  
G00-119-860"  
misc\_difference 8178  
/gene="EDH17B1"  
/note="C in clone lambda17-HSD5; T in clone lambda17-HSD3;  
G00-119-860"  
repeat\_unit 10880..11014  
/gene="EDH17B1"  
/note="left arm"  
/rpt\_family="Alu"  
/rpt\_type="direct"  
repeat\_unit complement(11393..11650)  
/rpt\_family="Alu"  
/rpt\_type="inverted"  
misc\_difference 1267  
/gene="EDH17B1"  
/note="C in clone lambda17-HSD5; T in clone lambda17-HSD3"  
repeat\_unit 13208..13930  
/gene="EDH17B1"  
/rpt\_family="Alu"  
/rpt\_type="direct"  
repeat\_unit 13520..13794  
/gene="EDH17B1"  
/rpt\_family="Alu"  
/rpt\_type="direct"  
repeat\_unit 13811..13913  
/gene="EDH17B1"  
/note="truncated"  
/rpt\_family="Alu"  
/rpt\_type="direct"  
gene 13914..20156  
/gene="EDH17B2"  
repeat\_unit 13914..20156  
/gene="EDH17B2"  
/note="G00-127-970"  
repeat\_unit complement(13914..14049)  
/rpt\_family="Alu"  
/rpt\_type="inverted"  
mRNA join(13999..15062,15156..15323,15471..15650,15826..15919,  
16436..16613,16700..17244)  
/gene="EDH17B2"  
/note="G00-127-970"  
prim\_transcript 13999..17244  
/gene="EDH17B2"  
/note="G00-127-970"  
exon 13999..15062  
/gene="EDH17B2"  
/note="G00-127-970"  
/number=1  
CAAT\_signal complement(14865..14868)  
protein\_bind /note="G00-127-970"  
/gene="EDH17B2"  
/note="G00-127-970"  
/bound\_moiety="transcription factor AP-2"

GC\_signal 14902..14915  
/gene="EDH17B2"  
TATA\_signal 14925..14931  
/note="G00-127-970"  
/gene="EDH17B2"  
CDS 14936..16613,16700..16969  
/note="G00-127-970"  
/join(14966..15062,15156..15323,15471..15650,15826..15919,  
16436..16613,16700..16969)  
/gene="EDH17B2"  
/EC number="1.1.1.62"  
/standard name="estradiol 17-beta-dehydrogenase"  
/codon\_start=1  
/product="17-beta-hydroxysteroid dehydrogenase"  
/protein\_id="AAB16942.1"  
/translation="MARTVLTGCGSSGIGLAVRLASDPSPKRYATLRLKTOG  
RLMEARALACPGPSLETJOLDVDRSKVAAAREVTLGRVYLVCNAGLGLPLEA  
LGEDAVASVLDVNVGTVRLCAFLPMKRRSGSRVLTGAVGLMGPENDVYCAK  
PALBGLCBRLAVLLPFGVHSLIECGVHTLPMKRYGSPREVDRLDITPHRFYQ  
YLAHKKQVREAAQNPESVAVETLALAPKTLTFTTERFLPLRLRLDDPGSNY  
VTAMREVFVGDVPAFAEAAGAAGAGGAGGAEAGSAVGDPELDPPAPQ"  
15063..15155  
/gene="EDH17B2"  
/note="G00-127-970"  
/number=1  
15156..15323  
/gene="EDH17B2"  
/note="G00-127-970"  
/number=2  
15324..15470  
/gene="EDH17B2"  
/note="G00-127-970"  
Intron  
Query Match 54.9%; Score 1414.6; DB 9; Length 21764;  
Best Local Similarity 99.7%; Pred. No. 1.7e-179;  
Matches 1428; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 1112 CATGACTGAGTGGATGACTGAGCTGTGTGCTGCTCAAGGCTGCTTCCAGACCA 1171  
DB 5056 CCGTTCACAGTGGATGACTGAGCTGTGTGCTGCTCAAGGCTGCTTCCAGACCA 5115  
QY 1172 GCGCAGTTCTGGGGGCGCGCCAGATCAAGGCTGTGCTGGAGCTGTGCGCG 1231  
DB 5116 GCGCAGTTCTGGGGGCGCGCCAGATCAAGGCTGTGCTGGAGCTGTGCGCG 5175  
QY 1232 CCTCTGTTTCTGACCTGTTTCTGAGGCAAGCTGTGTTTACCAGACTGCTCCTT 1291  
DB 5176 CCTCTGTTTCTGACCTGTTTCTGAGGCAAGCTGTGTTTACCAGACTGCTCCTT 5235  
QY 1292 CCAGGCGCAGCCCTCATCTGTGATGCTGCACAACTTTGGGGGAAACCATGCTTTT 1351  
DB 5236 CCAGGCGCAGCCCTCATCTGTGATGCTGCACAACTTTGGGGGAAACCATGCTTTT 5295  
QY 1352 TGGAGCCCTTAGAGCTGTGAACGAGGCCCAAGCTGCCCTCTTCCCACTCCAC 1411  
DB 5296 TGGAGCCCTTAGAGCTGTGAACGAGGCCCAAGCTGCCCTCTTCCCACTCCAC 5355  
QY 1412 CATGTAGGACAGGGGATGGCCCGAGGGGACATCAGCCAGAACGATGTATTTCCT 1471  
DB 5356 CATGTAGGACAGGGGATGGCCCGAGGGGACATCAGCCAGAACGATGTATTTCCT 5415  
QY 1472 CATGTAGGCTGGGGCTGGCGAAAGAACCACTGCGAATTTGGAGCTTGAGTACAG 1531  
DB 5416 CATGTAGGCTGGGGCTGGCGAAAGAACCACTGCGAATTTGGAGCTTGAGTACAG 5475  
QY 1532 CTTTGCGCGCGCGGATAGGGGCTCCCAACCGGACGAGGGGAGAGCTGAGGCTACT 1591  
DB 5476 CTTTGCGCGCGCGGATAGGGGCTCCCAACCGGACGAGGGGAGAGCTGAGGCTACT 5535  
QY 1592 GCTCCGAGTGTGTACAATGCTGCTCGGGAGGCTTGCAGAGGACCAATGTAAGCCGCT 1651  
DB 5536 GCTCCGAGTGTGTACAATGCTGCTCGGGAGGCTTGCAGAGGACCAATGTAAGCCGCT 5595

QY	1652	GGTACAGGCGGCGGCTCCCTCAACAGTGAATACCAAGCATCTGGTACAACCGATCTGAAGTGT	1771
Dp	5896	GGTACAGGCGGCGGCTCCCTCAACAGTGAATACCAAGCATCTGGTACAACCGATCTGAAGTGT	5655
QY	1712	TGAGGCGCTGAGCGGCTGCTCACAATCTGCTCCCTCCCTGGCCACACAGCCCGGCTTCG	1771
Dp	5656	TGAGGCGCTGAGCGGCTGCTCACAATCTGCTCCCTCCCTGGCCACACAGCCCGGCTTCG	5715
QY	1772	CTACGACCTGTGTGAACCTCACTCCGGCAGGCAAGTGAAGAGCTGGTCAAGCTTGTACTAAGA	1833
Dp	5716	CTACGACCTGTGTGAACCTCACTCCGGCAGGCAAGTGAAGAGCTGGTCAAGCTTGTACTAAGA	5779
QY	1832	GAAGGCAAGAAAGCGGCTACCTGAGCAAGAGCTGGCTCCCTGTTGAAGGCTGAGGCGT	1891
Dp	5776	GAAGGCAAGAAAGCGGCTACCTGAGCAAGAGCTGGCTCCCTGTTGAAGGCTGAGGCGT	5835
QY	1892	CCTGGCTATAGACTGCTGCTCCGGCACAAGTGAAGGCTGTGGCTAGTGAACAGCCGCTTCT	1951
Dp	5836	CCTGGCTATAGACTGCTGCTCCGGCACAAGTGAAGGCTGTGGCTAGTGAACAGCCGCTTCT	5895
QY	1952	GCTGGGCGAGCTGGGCTAGAGAGGCGGCGGACAGCGGCACTAGTAGGCGCGAGGCGGATTT	2011
Dp	5896	GCTGGGCGAGCTGGGCTAGAGAGGCGGCGGACAGCGGCACTAGTAGGCGCGAGGCGGATTT	5955
QY	2012	CTACAGCAGAAACAGCCGCTACCAAGCTACCTTGTGGGGGCGAGAGGCAACATCTTGA	2072
Dp	5955	CTACAGCAGAAACAGCCGCTACCAAGCTACCTTGTGGGGGCGAGAGGCAACATCTTGA	6014
QY	2072	CTATGACCACAAGAGAGCTGGGGGGTGTGTGGCCAACTATACACCTTCGTGGCGGCT	2131
Dp	6015	CTATGACCACAAGAGAGCTGGGGGGTGTGTGGCCAACTATACACCTTCGTGGCGGCT	6074
QY	2132	TTTCTCGTAGGCGCTGATTGAACAGTGTGGCCAGGAGCACTCCCTTCCAACAGACCAAT	2191
Dp	6075	TTTCTCGTAGGCGCTGATTGAACAGTGTGGCCAGGAGCACTCCCTTCCAACAGACCAAT	6134
QY	2192	TGACAAAAATGTTCTTCCAACTGGAGCAGAGCTTGTCTCAGCAAGCAGAGTACCCAG	2251
Dp	6135	TGACAAAAATGTTCTTCCAACTGGAGCAGAGCTTGTCTCAGCAAGCAGAGTACCCAG	6194
QY	2252	CCAGCGCGAGAGAGACATGTGGACCTTGGCCAAAGAAAGTCTTCCCAATATTACCCGG	2311
Dp	6195	CCAGCGCGAGAGAGACATGTGGACCTTGGCCAAAGAAAGTCTTCCCAATATTACCCGG	6255
QY	2312	CTGGGTGCGGCGCTTGTGTATAGATTGGCCACCACTGGGCGTTGTTCGCGTAATTC	2371
Dp	6255	CTGGGTGCGGCGCTTGTGTATAGATTGGCCACCACTGGGCGTTGTTCGCGTAATTC	6314
QY	2372	CAGGCGAATTCACAGGCGCCAGAGCTTGAACAGACATCAACAGATTAACCAAGGCTGGAG	2433
Dp	6315	CAGGCGAATTCACAGGCGCCAGAGCTTGAACAGACATCAACAGATTAACCAAGGCTGGAG	6374
QY	2432	GAGGCGCCACGCGCTGCTGTGTGGGCTGCACTGGGGGAGATTGAGGGGAAATGACTTGC	2491
Dp	6375	GAGGCGCCACGCGCTGCTGTGTGGGCTGCACTGGGGGAGATTGAGGGGAAATGACTTGC	6433
QY	2492	CTCCACACACACCCAAAGTGTGGATTTAAAGTACTTTTCTTCCACTTTAA	2544
Dp	6435	CTCCACACACACCCAAAGTGTGGATTTAAAGTACTTTTCTTCCACTTTAA	6487

RESULT	13
LOCUS	AF331668
DEFINITION	AF331668 2949 bp mRNA linear VRT 18-JUL-2001 Dromaius novaehollandiae lysosomal alpha-N-acetyl glucosaminidase mRNA, complete cds.
ACCESSION	AF331668
VERSION	AF331668.1 GI:14861377
KEYWORDS	.
SOURCE	Dromaius novaehollandiae (emu)
ORGANISM	Dromaius novaehollandiae Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Palaeognathae; Casuariiformes; Dromadidae;

REFERENCE	1 (bases 1 to 2949)
AUTHORS	Aronovich, E. L., Johnston, J. M., Wang, P., Gieser, U. and Whitley, C. B.
TITLE	Molecular basis of mucopolysaccharidosis type IIIB in emu ( <i>Dromaius novaehollandiae</i> ): an avian model of Sanfilippo syndrome type B
JOURNAL	Genomics 74 (3), 299-305 (2001)
MEDLINE	21309063
PUBMED	11414757
REFERENCE	2 (bases 1 to 2949)
AUTHORS	Aronovich, E. L. and Whitley, C. B.
TITLE	Direct Submission
JOURNAL	Submitted (21-DEC-2000) Peds & Inst. Human Genetics, University of Minnesota, 420 Delaware St. SE, Minneapolis, MN 55455, USA
FEATURES	Location/Qualifiers
source	1. 2949
	/organism="Dromaius novaehollandiae"
	/mol_type="mRNA"
	/db_xref="taxon:8790"
	13..2274
	/note="NAGLU"
	/cotton_start=1
	/product="lysosomal alpha-N-acetyl glucosaminidase"
	/protein_id="AAK73654.1"
	/db_xref="GI:14861378"
	/translation="MVAPVRLTALAAVAAAAPSPPLPPRAGAEADAPQAPVAVATLALRLIGPRAAVALSVDSGLAAGLDTPATRVHSPPGAAVAANVAGSGVAALAGLRHYRDLCCGHSMGSGROLRPLDPLPRVPAETRTAPRPGAAVAVAGTCTOSYSPAAWMDMAREEITDMMALSGINTLALAPAGQEAQVWRVYLSLGINSEIDETRTGAFPLANRMGLHGAAGPAPRAMHLKQLVQYRVLRNMSLMDITVLPFAAGHPPGVIRAFRVVATRLGGASHPECTVSTCYTLDPEDPMFOVIGTLFLKEIKSEFGVDHLYSAUTFEMNMLSSDPALESIRSAVFRSMSTGADRAVIMQGMVFOHPDPMQAPVALALHGVFLMGLVIDLEASBPVYQWTFSEFYGGPFIWMCMFEGNGLPCTVLAINHGPFAARFSPSTVGVGTGLVPGIEIQNDMIVYELMELGLRQPLDPSVAIVAAERYVGAAPAAASAMQLLSRSLYNCQGVVNNRNSPLVBRPLMDTETVWKNSDYVEAMRLLSGAELGSSPTFGYDLDVTRQAAQQLVSRYVLSIRGAPQSRSLPELLTGTGVLVYDILPELDGLSHRLPLGADTRESRAVATSPREAGQVEINARQVTLMDANKKGLVLDYGVPMWLPFSAVLESINSSSPFHQDPNQAVFQVERGFYNNKRRYPAPVGDITLISKKFLFKRYSPSMRRGRAGPA"
BASE COUNT	425 a 1025 c 1038 g 461 t
ORIGIN	
Query Match	43.3%; Score 1114.2; DB 5; Length 2949;
Best Local Similarity	69.6%; Pred. No. 2.8e-119;
Matches 1541; Conservative	0; Mismatches 668; Indels 6; Gaps 2;
QY	98 GACCATGAGAGCGGATGGCGGTGGCGCGCGCGGCGGTCTTCTTCTGCGCGGGGCCCGG 157
DB	33 GACGGCGCTGGCTCCATGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 92
QY	158 GGGGCGGGGAGCGAGCAGAGCGCGCGGGAGAGCGCGCGCGCGGTGGGGCGGTCTGTCGCCCGGCT 217
DB	93 GCGGGCGGGGCGCGAGAGACGCGCGCGCGAGAGAGCGCGCGGTGCGGGGCGCTGCGCGCGCGCT 152
QY	218 GCTGGGGGAGAGCGCCCGCGGCGCGAATTCTCCGTCCGTCCGTGGAGCGCGCTTGCGTGCCTAA 277
DB	153 GCTGGGGCGCGCGCGCGCGCGCGCGCTGTCGATGACGCGCTGCTGTCGTCGCGCGCGCGG 212
QY	278 GCGGGGCTTGGACACTAAGCGCTGGGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 337
DB	213 GCTGGAACCTTACCGGGTGCATCTGCGCGCGCGCGCGCGCGCGCGTGGCGCGCGCGCGCGCG 272
QY	338 CTCACAGGGGCTGGCGCGCGCGCGCGGGAGCTGCACCGCTACTCTGTCGCGAATTCTGTGTGCTG 397
DB	273 CTCACAGGGGCTGGCGCGCGCGCGCGCGCGCTTCACTGCGCGAATTCTGTGTGTGCTG 332
QY	398 CCAAGTGGGCTGGCGCGCGCGCTTCAAGCTGGCGCGCTGCGCGCGCACTGGCAGCGCGTGCAGG 457
DB	333 CCACTCTCTCTGCTGGAGAGGACAGCTTCGCTGCCGGAATCCCGTGCCTCCCGCGGATGCGCGC 392
QY	458 GGAAGTGAACGAGGCGCACGCGCAAGGTAACGCTATTATCCAGATATGTGTGACAGCAAG 517
DB	393 CGAATCCGCGCACCGCGCGCGCGCGAGGTTTCGCTACTACCAAGAAAGTCTGACACCGCAAG 452

QY	518	CTACTCCTTCGCTGGGTGGAGACTGGGCCCGCTGGAGAGCAAGATGACTGGATGGCGCT	577
Db	453	CTACTCCTTCGCTGGGTGGAGACTGGGCCCGCTGGAGAGCAAGATGACTGGATGGCGCT	512
QY	578	GATGGCATCAACTGGACCTGGACCTGGACCGGCCGAGAGGCATCTGGACCGAGTGA	637
Db	513	CAGCGGCATCAACTGGACCTGGACCTGGACCGGCCGAGAGGCATCTGGACCGAGTGA	572
QY	638	CGTGGCCTTGGGCGCTGACCGCAGGACGAGATTCATATAGTCTTTACTGCTCTGCTTCT	697
Db	573	CCTTCTCTTGGGCGCTGACCGCAGGACGAGATTCATATAGTCTTTACTGCTCTGCTTCT	632
QY	698	GGCCTGGGGCGAATGGGCACCTGACACTGGAGATGGCCCTTGCCCGCCCTCTGGCA	757
Db	633	GGCCTGGAGCCGATGGAGAACTCCACGGGCTGGGAGGGCGGCTGGCGCGGCTTGGA	692
QY	758	CATCAAGCAGCTTTACCTGACAGCACGGGTCTTGACACAGATGCGCTCTTGGCATGA	817
Db	693	CCTCAAGCAGCTTTACCTGACAGTACAGGGTCTTGAGAGAGTGGCTGCTGGTGGATAT	752
QY	818	CCCAATGCTGCTGCAATGGCGGGGCAATTTCCGAGGCTGTACACAGGGTGTCCCTCA	877
Db	753	CACGGTGTGCTGGGCTTGGACGGCCAGTGTCCCGAGGCGCTCTGAGGCTTTCCAG	812
QY	878	GGTCAATGTACAGAAAGTGGSCAGTTGGGGCCACTTTACCTGTCTACTCTGTCTCT	937
Db	813	TGTGATGTCACTGCGCTCTGGGGGCTGGAGCCACTTTGACTGCAAGTACTCTTGTACCTA	872
QY	938	CCCTTGCGTCCGGAAGACCCATATTCCTCCATCATCCGGAGACCTCTTCTGCGAGACT	997
Db	873	CTTGTCTGAACCCCGAGAGCCCATGTTTCAAGTCAATCGGACCTCTTCTTGAAAGACT	932
QY	998	GATCAAGAATTTGGCAACACACATCTATGGGGCCGACACTTTCATAGATGACACC	1057
Db	933	GATCAAGAGTTTGGCAACACACATCTATGAGGCTGACACTTTCATAGAGATGACCC	992
QY	1058	ACCTTCTTCAGAGCCCTCTACTTGGCCGACGCCACTGCTCGTCTATGAGGCGATGAC	1117
Db	993	GCTGTCTTCACACCCCGGCTACTCTTCAGGGTGAAGTGGCCGTCTTCAAGTCCATGAC	1052
QY	1118	TGCAGTGAATACTAGAGCTGTGTGGCTGTCCAAAGCTGGGCTTCTTCAGACACAGCCGA	1177
Db	1053	CGAGCGGACCCCAAGGACAGTGTGCTATGAGGGCTGGCTCTTTCACACACAGCCGA	1112
QY	1178	GTTCTGGGGGCCCGCCAGATCAAGGCTGTGTGGAAGCTGTGCCCGTGGCGGCTCT	1237
Db	1113	CTTCTGGGACGCGGCGAGTGTGGGGCGCTGTGTGACCGCGTGGCGCTTCGGGAGATGAT	1172
QY	1238	GGTTGTGAACGTTTGTGTGAAGCAACCTGTGTATACCCGACCTGCTCTTTCAGAGG	1297
Db	1173	CGTTTCAACCTTTCGCGAGATCCAGGCCGCTGTATCAAGTGAACGAGTCCCTTCAAGG	1232
QY	1298	CCAGCCCTTCACTTGGTACATGCTGACCAACTTTGGGGAAACCATGATCTTTTGGAGC	1357
Db	1233	GCAGCCCTTCACTTGGTACATGCTGACCAACTTTGGGGGCAACACCGGCTCTTGGGAC	1292
QY	1358	CCTAGAGGCTGTGAACGAGAGCCCAAGAGCTGCCGCTCTTTCCTCACTCACATGTGT	1417
Db	1293	CGTGAAGGCTATCAACACCGGCTCTTTCGACCGCGGCTTCCCACTCCACCATGTGT	1352
QY	1418	AGGCACGCGGATGGGCCCGAGAGGCTACGACAGAACAGATGGTCTATTCGCTCATGGC	1477
Db	1353	GGGACCGGGGTGTGTCCCGAGGGGATGAGACACACATGTGTATCAGAGCTGATGAA	1412
QY	1478	TGAGCTGGGCTGGCGAAGAGACCAATGCCAAGATTTGGCAAGCTTGGTGAACGATTGGC	1537
Db	1413	CGAATGGGCTGGCGCCAGAGAGCC---CCTGACCTCCCGAGCTGGGTGGCCCTCATATGC	1459
QY	1538	CGCCCGGCGATGGAGTCTCCACCCGAGACGACAGGGGCAACGAGAGGCTATGATCTCCG	1597
Db	1470	GGAGGCGCGCTACGAGAGCCCGAAGCGTGCACGGCCAGCGCTGTGGAGCTGCTCTCCG	1529
QY	1598	GAGTGTGACACTGCTCCGGGAGGGCTTGACGGGGCCACCAATGTGATCCGCTGTGACG	1657

Db	1530	CAGCGTGTACAACTGACAC--GGGGGCTGTGGCTCAACCAACACCCGAGCCCTGATGCG	1581
QY	1658	GGGGCGGCTCCCTTACAGATGAATATACAGCATCTGAGTACAACCGATCTGATGTGTTGAGGC	1711
Db	1587	CCGGCCCTCCCTGGCCCATGAGAACCCGAGAGTCTGTATCAACAAGACGACGTATAGAGGC	1641
QY	1718	CTGGGGGCTGTGCTGTCTCAATCTGTCTCCCTCCCTGGCCACACAGCCCGCTTCCGTACGA	1771
Db	1647	CTGGGGCTGTGCTGTGAGCCGCGCGCGAGAGTGGAGCTTCCAGCCCACTTCCGTACGA	1700
QY	1778	CTGTGTGACCTCACTCTCGGACAGGACAGTGCAGAGAGTGTGACGTTGTACTATGAGAGGC	1831
Db	1707	CCTGGCGGACGTCAACGCGGACAGCGGCCACGACAGTGTGTAGCCAGTACTACCTAGCAT	1766
QY	1838	AAGAAAGCCTTACCTTGAGCAAGAGAGCTGGCCTCCTCTTGAAGGCTGAGAGCGTCTGAC	1891
Db	1767	CCGGGACGCTTCCACAGACCCGCTCGCTCCGAGAGCTGTCTCAAGCGCGGCGGTGTCTGT	1822
QY	1898	CTATGAGGCTGCTGCGCGGACCTGAGACAGAGTCTGTGCTGTGACAGCCGCTTCTTGAGG	1951
Db	1827	CTACGACCTGTCTCCCGAGCTGTGACAGGCTCTCTGTCCAGCCACGCGCTTCTCTGTGAGG	1888
QY	1958	CAGCTGTCTAGAGCAGGCGCCGAGACCGGACATGATGAGGCGCGAGCCGATTTTACGA	2011
Db	1887	CCGCTGTGTGAGAGACCGCCCGTGGCGTGGCCACACAGGACAGGGAGGCGAGCACTAGCA	1944
QY	2018	GCAAGAACGCGGCTTACCACTGACCTTGTGGGGGCGAAGAGCAATCTTGGACTATGC	2077
Db	1947	GCTGAACCGCCGGAACAAGAGTGAAGCTGTGGGGGGCCCAACGCAATCTTGGATTAGCC	2006
QY	2078	CAACAGAGAGCTGACGCGGGGTGTGTGGCCAACTACTACACCCCTGCTGGGCGCTTTCT	2131
Db	2007	CAACAGAGAGCTGGGGGGGCTGT	2066
QY	2138	GAGGCGCTGT	2197
Db	2067	CTGTGCTTGT	2128
QY	2198	AAATGTCTTCAACTGTGAGCAGGCGCTTGTCTCAGACAGAGAGGTATCCCAAGCAGCC	2251
Db	2127	GGCGCTCTTCCAGGTATGAGAGGCTTCTATTTACACAGAGAGGCTATCCCAAGCAGCC	2188
QY	2258	GCGAGAGACACTGTGTGACCTGTGGCCAGAGAGATCTTCTCAATATTTACCCCGG	2312
Db	2187	TGTGGGGGACACGCTGTGAGATCTCAAGAGATATTTCTCAATATTTACCCCGG	2241
RESULT 14			
AF003255			
LOCUS	AF003255	7721 bp	DNA linear ROD 13-DEC-1999
DEFINITION	Mus musculus alpha-N-acetylglucosaminidase gene, complete cds.		
ACCESSION	AF003255		
VERSION	AF003255.1	GI:3329360	
KEYWORDS			
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 7721)		
AUTHORS	Li, H.H., Yu, W.H., Rozenburg, N., Zhao, H.Z., Lyons, K.M., Anagnostaras, S., Farnsworth, M.S., Suzuki, K., Vanier, M.T., and Neufeld, E.F.		
TITLE	Mouse model of Sanfilippo syndrome type B produced by targeted disruption of the gene encoding alpha-N-acetylglucosaminidase		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (25), 14505-14510 (1999)		
MEDLINE	20056274		
PUBMED	10588735		
REFERENCE	2 (bases 1 to 7721)		
AUTHORS	Li, H.H., Zhao, K.W., and Neufeld, E.F.		
TITLE	Cloning and characterization of gene and cDNA encoding mouse alpha-N-acetylglucosaminidase		
JOURNAL	Unpublished		

REFERENCE 3 (bases 1 to 7721)  
 AUTHORS Li, H.H., Zhao, K.W. and Neufeld, E.F.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-MAY-1997) Biol. Chem., UCLA, 10833 Le Conte Ave., Los Angeles, CA 90095-1737, USA  
 FEATURES  
 source Location/Qualifiers  
 1..7721  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="129sv"  
 /db\_xref="taxon:10090"  
 /chromosome="11"  
 join(206..582,1160..1307,1966..2112,2287..2372,3918..4174,6311..7515)  
 /product="alpha-N-acetylglucosaminidase"  
 join(206..582,1160..1307,1966..2112,2287..2372,3918..4174,6311..7515)  
 /codon\_start=1  
 /product="alpha-N-acetylglucosaminidase"  
 /protein\_id="AAC28842.1"  
 /db\_xref="GI:3129361"

mrna  
 join(206..582,1160..1307,1966..2112,2287..2372,3918..4174,6311..7515)  
 /product="alpha-N-acetylglucosaminidase"  
 join(206..582,1160..1307,1966..2112,2287..2372,3918..4174,6311..7515)  
 /codon\_start=1  
 /product="alpha-N-acetylglucosaminidase"  
 /protein\_id="AAC28842.1"  
 /db\_xref="GI:3129361"

BASE COUNT 1735 a 2050 c 2056 g 1880 t  
 ORIGIN

Query Match 34.0% Score 874.8; DB 10; Length 7721;  
 Best Local Similarity 82.1%; Pred. No. 1.7e-107;  
 Matches 1021; Conservative 0; Mismatches 217; Indels 6; Gaps 1;  
 1116 ACTGAGTGAATCTAGAGCTGTGTGCTGCTCCAGAGCTGCTCTTCCAGACACAGCCG 1175  
 6304 ACTGATAGTGAACCTGATGCTGTGCTGCTCCAGAGCTGCTCTTCCAGACACAGCC 6363  
 1176 CAGTCTGGGGGCCCCCAATCAAGGCTGTGCTGGAAGCTGCTGCTGCTGCTGCTG 1235  
 6364 CAGTCTGGGGGCCCCCAATCAAGGCTGTGCTGGAAGCTGCTGCTGCTGCTGCTG 6423  
 1236 CTGTTCTGAGACCTGTTGCTGAGAGCAGCTGTGATACCCGACAGCTCTCTCCAG 1295  
 6424 CTGTTCTGAGACCTGTTGCTGAGAGCAGCTGTGATACCCGACAGCTCTCTCCAG 6483  
 1296 GGCACGCTCTCATCTGTGATGCTGCAACAATTGGGGGAAACATGATCTTTTGA 1355  
 6484 GGCACGCTCTCATCTGTGATGCTGCAACAATTGGGGGAAACATGATCTTTTGA 6543  
 1356 GGCCTAGAGAGCTGTGAACGAGAGCAGAGCTGCTCTCTTCCCAATCCACATG 1415  
 6544 GGCCTAGAGAGCTGTGAACGAGAGCAGAGCTGCTCTCTTCCCAATCCACATG 6603  
 1416 GTAGGACGAGGATGCTGCTGATGCTGCAACAATTGGGGGAAACATGATCTTTTGA 1475  
 6604 GTAGGACGAGGATGCTGCTGATGCTGCAACAATTGGGGGAAACATGATCTTTTGA 6663  
 1476 GTAGGAGCTGCTGAGAAAGGACCAAGTGCAGATTGGAGAGCTGCTGCTGCTGCTG 1535  
 6664 GTAGGAGCTGCTGAGAAAGGACCAAGTGCAGATTGGAGAGCTGCTGCTGCTGCTG 6723  
 1536 GCGCGCCGCGGATAGGGGTCTCCACCGAGCGAGAGGCGAGCGCTGAGAGCTAGCTC 1595  
 6724 GCGATCCGCGGATAGGGGTCTCCACCGAGCGAGAGGCGAGCGCTGAGAGCTAGCTC 6783

1596 CCGAGTGTGTACACTGCTCCGCGGAGAGCTGACAGGCGCACATGTAACCCGCTGCTC 1655  
 6784 AGAAGTGTATCAACTGCTCTGGGAGGCGGTGACATGGGCAACAATGAAAGCCCTTGGTC 6843  
 1656 AGCGCGCGCTCCCTCAAGATGATACAGACATCTGTTCAACCCATCGATGTTGAG 1715  
 6844 AAGCGCGCGCTCCCTCAAGATGATACAGACATCTGTTCAACCCATCGATGTTGAG 6903  
 1716 GCGTGGCGGCTGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1775  
 6904 GCTTGGCGACTGCTGTTAAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6963  
 1776 GACTGCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1835  
 6964 GACTGCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7023  
 1836 GCAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1895  
 7024 GCAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7083  
 1896 GCGTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1955-  
 7084 GCTTATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7143  
 1956 GCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2015  
 7144 GGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2203  
 2016 GAGCAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2075  
 7204 GAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7263  
 2076 GCGAAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2135  
 7264 GCGAAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7323  
 2136 CTGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2195  
 7324 TTGGGAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7383  
 2196 AAAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2255  
 7384 AAGAAAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7443  
 2256 CCGAGAGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2315  
 7444 CCGAGAGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7497  
 2316 GTGGCGGCTCTTGTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2355  
 7498 CAGCTGACTCTTGTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7541

RESULT 15  
 AF363242 10938 bp DNA linear ROD 02-MAY-2002  
 LOCUS AF363242  
 DEFINITION Mus musculus N-acetylglucosaminidase (Naglu) and 17-beta-hydroxysteroid dehydrogenase (Hsd17b1) genes, complete cds.  
 ACCESSION AF363242  
 VERSION AF363242.1 GI:20385159  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 10938)  
 Gloeckner, C.J., Breiling, R., Moeller, G. and Adamski, J.  
 Characterization of HSD17B1 gene in mice  
 Unpublished  
 2 (bases 1 to 10938)  
 Gloeckner, C.J., Breiling, R., Moeller, G. and Adamski, J.  
 Direct Submission  
 JOURNAL Submitted (23-MAR-2001) Institute for Experimental Genetics,

GSF-National Research Center for Environment and Health,  
Ingolstaedter Landstr. 1, Neuherberg 85764, Germany

## FEATURES

**Source**

1. 10938

/organism="Mus musculus"

```
/lib_type="genomic DNA"
```

```
/db xref="taxon:10090"
```

```
/sex="female"
```

```
/tissue_type="spleen"
```

<206. .>7515

```

/gene="Naglu"
join(<206 582 1160 1

```

JOHNSON, J. 1962, 1180. .1  
3918, .4174, 6311, .7515

Db

7498 CAGCCTGACTCTTTGTGTGACAGATTAGCCATGCGAGGACCTGCT 7541

Search completed: February 13, 2004, 12:03:30  
Job time : 9404 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 13, 2004, 07:06:01 ; Search time 672 Seconds

(without alignments)  
10343.829 Million cell updates/sec

Title: US-09-836-613-1

Sequence: 1 CCCGGCGCTTAGCCCTCGGGT.....AGTCGAGCGCGCGCAATTC 2575

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

N\_Geneseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2575	100.0	2575	18	AAT67163
2	1426.6	55.4	10380	18	AAT67163
3	369.4	14.3	2661	23	AB122605
4	369.4	14.3	4740	23	AB122604
5	305.2	11.9	706	24	ABQ32798
6	305.2	11.9	706	24	ABQ32799
7	304.2	11.8	710	24	ABQ14172
8	304.2	11.8	710	24	ABQ14173

c	9	259.8	10.1	706	24	ABQ32800
c	10	259.8	10.1	706	24	ABQ32801
c	11	257.2	10.0	710	24	ABQ14170
c	12	257.2	10.0	710	24	ABQ14171
c	13	89.2	3.5	7222	20	AB122584
c	14	80.4	3.1	114955	20	AAK53492
c	15	71	2.8	4020	18	AA151361
c	16	70.4	2.7	5173	18	AAT89783
c	17	70.4	2.7	5173	24	ABN96859
c	18	70.4	2.7	5175	23	ABV22988
c	19	70.4	2.7	5175	23	ABV28823
c	20	70	2.7	10144	24	ABN95670
c	21	70	2.7	114955	20	AAK53492
c	22	68.6	2.7	2561	22	AAH26500
c	23	68.2	2.6	2271	24	ABZ35021
c	24	68.2	2.6	5181	24	AB168935
c	25	66.2	2.6	3198	20	AAK02974
c	26	64.4	2.5	1614	22	AAH26499
c	27	64.4	2.5	3957	22	ABV55290
c	28	64.4	2.5	4225	25	ABV55290
c	29	64.4	2.5	14225	22	AAH26495
c	30	64.4	2.5	154746	24	AAD25519
c	31	63.6	2.5	5811	24	ABN97455
c	32	63.2	2.5	43058	24	ABN97455
c	33	63.2	2.5	43058	24	AB164982
c	34	63.2	2.5	43058	24	AB164982
c	35	63.2	2.5	43058	24	AB164982
c	36	62.8	2.4	15672	12	AAQ10613
c	37	62.4	2.4	3957	22	AAK09686
c	38	62.2	2.4	4411529	22	AA199682
c	39	61.4	2.4	1562	22	AAH48175
c	40	61.2	2.4	24379	18	AAT93095
c	41	61.2	2.4	24379	19	AAV5925
c	42	61	2.4	15355	22	AAH57405
c	43	61	2.4	15359	22	ABH08724
c	44	61	2.4	15359	23	AAK59156
c	45	61	2.4	4403765	22	AA199683

## ALIGNMENTS

RESULT 1	
ID	AAT67163 standard; cDNA; 2575 BP.
AC	AAT67163;
XX	
DT	20-AUG-1997 (first entry)
XX	
DE	Human alpha-N-acetylglucosaminidase cDNA.
XX	
KM	Alpha-N-acetylglucosaminidase; mucopolysaccharidosis type IIIB;
KW	gene therapy; enzyme replacement therapy; diagnosis; ss.
XX	
OS	Homo sapiens.
XX	
PH	Key
FT	CDS
FT	sig_peptide
FT	mat_peptide
FT	
XX	
PN	WO9719177-A1.
XX	
PD	29-MAY-1997.
XX	
PF	22-NOV-1996;
XX	
PR	96WO-AU00747.
XX	
XX	23-NOV-1995;
XX	95AU-0006748.

Oligonucleotide fo  
Oligonucleotide fo  
Oligonucleotide fo  
Oligonucleotide fo  
Drosophila melanog  
Human adenovirus A1  
orf virus genomic  
Human immunodefici  
Gene #357 used to  
Human prostate exp  
Human prostate exp  
Gene #2168 used to  
Human adenovirus A1  
Rabbit low density  
Human gene express  
Kidney cancer rela  
Human IL-1ra BAC c  
Human low density  
HSV-2 immediate ea  
Mouse SGA2 polypep  
Human low density  
Human herpesvirus  
Human herpesvirus  
S. macromycetis  
Gene #3953 used to  
Lung cancer relate  
Lung cancer relate  
Ritandin receptor  
HSV-2 immediate ea  
Mycobacterium tube  
Rice glaberrima 2  
Streptomyces fredo  
Streptomyces roseo  
Human skeletal mus  
Human tyandine re  
DNA encoding novel  
Mycobacterium tube

PA (WOMEN) WOMEN'S & CHILDREN'S HOSPITAL.  
 XX Anson DS, Blanch L, Hopwood JJ, Scott H, Weber B;  
 PI MPI: 1997-298114/27.  
 XX P-PSDB; AAM18017.  
 DR  
 XX Nucleic acid encoding mammalian alpha-N-acetylglucosaminidase -  
 PT used for the diagnosis and treatment of mucopolysaccharidosis type  
 PI IIIB, also used in gene therapy  
 XX  
 PS Claim 7; Page 46-50; 79pp; English.  
 CC A cDNA clone (AAT67163) codes for human alpha-N-acetylglucosaminidase  
 CC (AAM18017), an enzyme that can hydrolyse the terminal alpha-N-  
 CC acetylglucosamine residues at the non-reducing terminus of fragments  
 CC of heparan sulphate and heparin. It was isolated from a human  
 CC peripheral blood leukocyte cDNA library using probes based on  
 CC isolated peptide fragments (see also AAM18018-20) of the enzyme and  
 CC by examination of the corresponding genomic gene (see also AAT67164).  
 CC The isolated cDNA, and primers/probes based on the sequence or on  
 CC or its complementary strand, can be used to investigate, diagnose  
 CC and treat alpha-N-acetylglucosaminidase deficiency, such as  
 CC patients suffering from mucopolysaccharidosis type IIIB.  
 CC Administration is by oral, i.v., i.p., enzyme replacement therapy,  
 CC gene therapy or other routes.  
 CC  
 XX  
 SQ Sequence 2575 BP; 440 A; 819 C; 833 G; 483 T; 0 other;  
 Query Match 100.0%; Score 2575; DB 18; Length 2575;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCGGGCTTACCTTCGGGTCACAGTGGCCGAGCCGCGACGCTGATTGACCGCGGCG 60  
 Db 1 CCCGGGCTTACCTTCGGGTCACAGTGGCCGAGCCGCGACGCTGATTGACCGCGGCG 60  
 QY 61 CCCGACCCCTTGGCGCGTGGCGGAGCCGCGACGAGCTGAGACCATGAGGCGGTTG 120  
 Db 61 CCCGACCCCTTGGCGCGTGGCGGAGCCGCGACGAGCTGAGACCATGAGGCGGTTG 120  
 QY 121 CCGCGGCGGTTGGGGGCTCTTCTCTGCGCGGAGCGCGGAGCGCGAGCGAGCGCC 180  
 Db 121 CCGCGGCGGTTGGGGGCTCTTCTCTGCGCGGAGCGCGGAGCGCGAGCGAGCGCC 180  
 QY 121 CCGCGGCGGTTGGGGGCTCTTCTCTGCGCGGAGCGCGGAGCGCGAGCGAGCGCC 180  
 Db 121 CCGCGGCGGTTGGGGGCTCTTCTCTGCGCGGAGCGCGGAGCGCGAGCGAGCGCC 180  
 QY 181 GGGAGGCGGCGCGCTCGCGCGGCTCGTGGCGCGGCGCTGCGGCGGCGGCGCGCG 240  
 Db 181 GGGAGGCGGCGCGCTCGCGCGGCTCGTGGCGCGGCGCTGCGGCGGCGGCGCGCG 240  
 QY 241 ACTTCTCCGTTGCGGTGAGCGCGCTCTGCTGCGCAAGCCGCGGCTTGAACCTACAGCC 300  
 Db 241 ACTTCTCCGTTGCGGTGAGCGCGCTCTGCTGCGCAAGCCGCGGCTTGAACCTACAGCC 300  
 QY 301 TGGGCGGCGGCGCGCGCGCGCGCTGCGGCGCGCGCTGCGGCGGCGCGCGCG 360  
 Db 301 TGGGCGGCGGCGCGCGCGCGCGCTGCGGCGCGCGCTGCGGCGGCGCGCGCGCG 360  
 QY 361 CGGCGGCTGACCGCTTACTGCGGCACTTCTGTGCTGCCACGTCGCTGCTGCTGCTC 420  
 Db 361 CGGCGGCTGACCGCTTACTGCGGCACTTCTGTGCTGCCACGTCGCTGCTGCTGCTC 420  
 QY 421 AGCTGCGCTTGCCTGCGGCACTGCGCAAGCGCTGCGGCGGAGCTGACCGAGCGCA 480  
 Db 421 AGCTGCGCTTGCCTGCGGCACTGCGCAAGCGCTGCGGCGGAGCTGACCGAGCGCA 480  
 QY 481 ACAGGTACCGCTATTACAGAAATGTGACGCAAGAACTACTCTCTTGTGTGTGGAACT 540  
 Db 481 ACAGGTACCGCTATTACAGAAATGTGACGCAAGAACTACTCTCTTGTGTGTGGAACT 540  
 QY 541 GGGCGCGCTGGAGCGAGATAGACTGATGGCGCTGAATGGCATCACTGGCACTGG 600  
 Db 541 GGGCGCGCTGGAGCGAGATAGACTGATGGCGCTGAATGGCATCACTGGCACTGG 600

QY 601 CCGAGCGGCGGAGGAGCCATCTGAGCGGGGTGACCTGGCTTGGGCTGACCCAG 660  
 Db 601 CCGAGCGGCGGAGGAGCCATCTGAGCGGGGTGACCTGGCTTGGGCTGACCCAG 660  
 QY 661 CAGAGTCAATAGTCTTCTTACTGCTGCTCTTCTGCGCTCGGCGCGAATGGCAAC 720  
 Db 661 CAGAGTCAATAGTCTTCTTACTGCTGCTCTTCTGCGCTCGGCGCGAATGGCAAC 720  
 QY 721 TGCACACCTGGAGTGGCCCGCCCTCTGCGCACATCAAGAGCTTACTCTGAGC 780  
 Db 721 TGCACACCTGGAGTGGCCCGCCCTCTGCGCACATCAAGAGCTTACTCTGAGC 780  
 QY 781 ACCGGTCTGAGCCAGATGCGCTCTTGGCATGACCCAGTGTGCTGCTGATTCGCG 840  
 Db 781 ACCGGTCTGAGCCAGATGCGCTCTTGGCATGACCCAGTGTGCTGCTGATTCGCG 840  
 QY 841 GGCATGTTCCGAGGCTGACAGAGGTGTTCCCAAGTCAATGTCAGAAAGTGGCA 900  
 Db 841 GGCATGTTCCGAGGCTGACAGAGGTGTTCCCAAGTCAATGTCAGAAAGTGGCA 900  
 QY 901 GTTGGGCGCATTTAATGTTCTTACTGCTGCTCTTCTGCTCTGCGAGAGACCCCA 960  
 Db 901 GTTGGGCGCATTTAATGTTCTTACTGCTGCTCTTCTGCTCTGCGAGAGACCCCA 960  
 QY 961 TATTCCTCATATCGGAGGCTCTTCTGCGAGAGCTGATCAAGAGTTTGGACAGACC 1020  
 Db 961 TATTCCTCATATCGGAGGCTCTTCTGCGAGAGCTGATCAAGAGTTTGGACAGACC 1020  
 QY 1021 ACATCTATGGGCGGACCATCTTCAATGAGATGACAGCCACTTCTCAGAGCCCTCTAC 1080  
 Db 1021 ACATCTATGGGCGGACCATCTTCAATGAGATGACAGCCACTTCTCAGAGCCCTCTAC 1080  
 QY 1081 TTGCGGACACCAACCATGCGCTTATAGAGCCATGATCTGAGTGTATCTAGGCTGTGT 1140  
 Db 1081 TTGCGGACACCAACCATGCGCTTATAGAGCCATGATCTGAGTGTATCTAGGCTGTGT 1140  
 QY 1141 GCGTGTCCAAAGCTGAGGCTCTTCCAGACAGCCGAGTTCGAGGCGCGCGCGAGATCA 1200  
 Db 1141 GCGTGTCCAAAGCTGAGGCTCTTCCAGACAGCCGAGTTCGAGGCGCGCGCGAGATCA 1200  
 QY 1201 GGGCTGTGTGGAGCTGTGCGCGCGGCGGCGCTCTGCTGAGCCTGTTTGTGAGA 1260  
 Db 1201 GGGCTGTGTGGAGCTGTGCGCGCGGCGGCGCTCTGCTGAGCCTGTTTGTGAGA 1260  
 QY 1261 GCGAGCTGTGTATACCGGCACTGCTCTTCCAGAGGCGCAACCTTCACTGTGTGAGC 1320  
 Db 1261 GCGAGCTGTGTATACCGGCACTGCTCTTCCAGAGGCGCAACCTTCACTGTGTGAGC 1320  
 QY 1321 TGCACAACTTTGGGGGAAACCATGCTCTTTTGGAGCCCTAGAGGCTGTGAACGAGGCC 1380  
 Db 1321 TGCACAACTTTGGGGGAAACCATGCTCTTTTGGAGCCCTAGAGGCTGTGAACGAGGCC 1380  
 QY 1381 CAGAGCTGTGCGGCTCTTCCCACTGACATGTTAGGCAAGGAGATGGCCCCAGAG 1440  
 Db 1381 CAGAGCTGTGCGGCTCTTCCCACTGACATGTTAGGCAAGGAGATGGCCCCAGAG 1440  
 QY 1441 GCATCAGCAAGAACAGATGTCTATTCCTCATGAGCTGAGCTGAGCGAGAAAGAGCC 1500  
 Db 1441 GCATCAGCAAGAACAGATGTCTATTCCTCATGAGCTGAGCTGAGCGAGAAAGAGCC 1500  
 QY 1501 CAGTCCAGATTTGGCAGCTGCGGAGCGAGCTTGGCGCGCGGCGGATGAGGAGCTTCC 1560  
 Db 1501 CAGTCCAGATTTGGCAGCTGCGGAGCGAGCTTGGCGCGCGGCGGATGAGGAGCTTCC 1560  
 QY 1561 ACCGGAAGCGAGGCGAGCGTGAAGGCTACTGCTCCGAGAGTGTGACAACTGCTCCGG 1620  
 Db 1561 ACCGGAAGCGAGGCGAGCGTGAAGGCTACTGCTCCGAGAGTGTGACAACTGCTCCGG 1620  
 QY 1621 AGGCTGAGAGGCGCAATGCTAGCCGCTGTGAGGCGGCGCTCTTACAGATGAATA 1680  
 Db 1621 AGGCTGAGAGGCGCAATGCTAGCCGCTGTGAGGCGGCGCTCTTACAGATGAATA 1680  
 QY 1681 CCAGATCTGGTACAAACCATGTGATGTGTGAGGCGCTGCGGCTGTGCTACATCTG 1740





CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2661 BP; 657 A; 653 C; 698 G; 653 T; 0 other;

Query Match 14.3%; Score 369.4; DB 23; Length 2661;  
Best Local Similarity 54.0%; Pred. No. 8.2e-57;  
Matches 901; Conservative 0; Mismatches 741; Indels 27; Gaps 6;

```

QY 491 CTATTACCAAGATATGTGACACGCAAGCTACTCTCTGATGTGTGGAGTGGCCGCTG 550
DB 774 CTATACCAAGATATGTGACACGCAAGCTACTCTCTGATGTGTGGAGTGGCCGCTG 833
QY 551 GGAAGCAAGATATGTGACACGCAAGCTACTCTCTGATGTGTGGAGTGGCCGCTG 610
DB 834 GCGGCTGACCTGATGTGATGAGCTGATGAGGATAGTCTATATGCTC--GCCGAT 890
QY 611 CCAGAGAGCCATCTGGACGCGGTGTACTGTGCTTGGGCTGACCCAGCAGAGATCAA 670
DB 891 CCAGAGAGCCATCTGGATGAGGTGTACGAGATATGGGCTTCAGATGAGGAATTCGA 950
QY 671 TGAATTTTAACTGATGCTGCTCTCTGAGCTGGGCGGAATGGGCACTGTGACCTG 730
DB 951 CGAGCATCTGCTGAGACCACTTTTCAGGCTGCGACGCGCATGCGCAATACGCGGATG 1010
QY 731 GGAATGCCCCCTGCCCCCTCTCTGACATCAAGCAGCTTTACCTGACGACCGGCTCT 790
DB 1011 GCGGAGCACTGACACCGGATGCGGACGCTATCAATTACTCTCTGACGAGAGATTTAT 1070
QY 791 GGACCAAGATGCGCTCTCTGCGCATGACCCAGTGTGCTGCTGATTCGCGGCGATGTC 850
DB 1071 CACTGCCCAAGCATCTGGGCAATGAGCGTGTGCGCTTGGCGCTTGTGATGTGCC 1130
QY 851 CGAGGCTGTACCAAGGATTTCCCTCAGATGATCAAGAGATGGGAGATTTGGGCGCA 910
DB 1131 ACCTGCTCTGAAAGCTGTGATCCGATGATGATGATGAGATGACGATGAGACCA 1190
QY 911 CTTTAACTGCTTCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 970
DB 1191 GTTCCCGCATGACATGCTGCTGCGGTTATTTGTGAACCAAGCAATCTTTTGAAGA 1250
QY 971 CATGGAGAGCTTCTCTGAGAGCTGATCAAGAGTTTGGACAGACCACTCATG 1030
DB 1251 GATAGCTGCGCTTTTGTGATATATCATACGAATATGTTGCAATCAATCTTCTT 1310
QY 1031 GCGCGACACTTTCAATGATGAGCAGCCACTTCTCTGAGACCTTCTCTCTCTCTCTCT 1090
DB 1311 CTGCGATCCCTTCAAGCACTGGAGCCACCGGTAGCAAGCCGAATATAGCTTTAAC 1370
QY 1091 CACCACTGCGCTATGAGAGCCATGACTGATGATCTGAGCTGTGTGCTGTCCA 1150
DB 1371 GCGGCGGCGATTTATGAATCTATGCGGCAATGATCTCTGAGCCATTTGTTCTTCA 1430
QY 1151 AGGCTGCTCTTTCACACACAGCCGAGTTCTGGGAGCCCGCCAGATCAGGGCTGTGCT 1210
DB 1431 GGGCTGATGTTGTGTAAGATTC--CTTTGACACAGCAATGCTGAAAGCTTTTCT 1487
QY 1211 GGAAGCTGTGCCCCGTGCGGCTCTCTGATGTGACCTGTGTTGTGAGAGCCACTGT 1270
DB 1488 CACGCGCGACACCTGTGCTGCTTTTGTGCTCTGATGACAGAGCAACAGTTTCCGA 1547
QY 1271 GTATACCGCAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1330
DB 1548 ATACAGATTAAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1607
QY 1331 TGGGGAACCATGATCTTTTGTGAGCCCTAGAGGCTGTGAACGAGGCGCCAGAACTGC 1390
DB 1608 CCGCGGACACTTGAATGTTTGAATCAGCAAGCTGATTAATTCGGAAATGAAAGAGC 1667
QY 1391 CCGGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1450
DB 1668 GCGTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1727
QY 1451 GAACGAATGTCTATTTCCCTCATGCTGAGCTGGGCTGGGAAAGAACCCAGTCCAGA 1510

```

```

DB 1728 GAACATATGATGATTTGCTTACCTTGGAGCGCGCTGAGATA--ATACATCCCTAGA 1784
QY 1511 TTGGACGCTGGGAGACAGCTTTGCGCCCGCGATGAGGCTCTCCACCCGAGAGC 1570
DB 1785 TCTGATAGCTGTGTTTACGAATTTCTCACACAGCCGATGAGCTGTAAGATGAGCGTTT 1844
QY 1571 AGGGCAAGCTGAGAGCTACTGCTCCGAGTGTGTACAATCTCCGAGAGCGCTGAG 1630
DB 1845 GGAGCAGGCGTGTGCTCTCTAAAGAAATAGCGTCTATTCCTCCGTGTGACAGAAAT 1904
QY 1631 GGGCCCAATGCTAGCCGCTGCTGACGCGCGCTGCTCTCAAGATGAATACAGCATGTG 1690
DB 1905 GCGGGAAGATGATGATGAGCGCA--GGCTTCTTCAACCAAGAACCTTTCATCTG 1961
QY 1691 GTACACCGATCTGATGTGTTTGAAGCCTGAGGCTGTGCTGCTCATCTGCTCTCTCT 1750
DB 1962 GTACATGCAATGATGAGTGTGATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2021
QY 1751 GGC-----CACCAAGCCCCCTTCCGCTACGACTGCTGACTCTGCGCA 1798
DB 2022 ACCGCTAGAGATTAACATACAGATATACGAACAGATTTGGTGAATACATACAGCA 2081
QY 1799 GGCAGTGCAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1856
DB 2082 ATTCTGCAATCAAGCGCAATCACTGTTATATCAATCAATGATGAGCTGCTA--TAGAAA 2138
QY 1859 GAGCTGCTCTCTCTGTTGAGGCTGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1918
DB 2139 GCGACAGGTTTGGGTTTCAATTCCTTAAAGCTGTAAGCTGTTGAACTGTTGATGACAT 2198
QY 1919 GAGCAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1978
DB 2199 GAGCTGATCTTGTGAGCAGCAGCAAAATTTCTCTGCGCACTGCTGCGAGCGCAA 2258
QY 1979 AGCAGCGCACTGATGAGAGCGCGAGCTTCTCAAGAGCAAGACCGCTACAGCT 2038
DB 2259 GCAAGCGCAACGAACTGCGCAAAAGAAATTTGAAATTCGATGAGGAAACCAAT 2318
QY 2039 GACCTTGTGGGCGCAAGAGCAATCTCTGATATGCCAACAAGAGCTGGGCGGCTT 2098
DB 2319 TACGCGCTGGGCTCTGATGCTGATCTGATGCTGATGCTGATGCTGATGCTGATGCT 2378
QY 2099 GGTGCGCAACTACACCCCTGCTGCGGCTTTTCTGAGAGCGCTG 2147
DB 2379 GTTGAAGGATCTATGAGCTGTGAGAGCTTCTGGAAGATGTG 2427

```

#### RESULT 4

ABL22604/C  
ID ABL22604 standard; DNA; 4740 BP.

XX ABL22604;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 19285.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001MO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PMD, Myers EW;  
 XX MPI; 2001-65860/75.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 XX interactions -  
 XX  
 XX Claim 1; SEQ ID NO 19285; 21bp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
 CC sequences (AB101840-AB116175) and the encoded proteins  
 CC (AB157737-AB172072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WFO  
 CC at ftp.wipo.int/pub/published\_pcc\_sequences.  
 CC  
 XX  
 XX  
 XX Sequence 4740 BP; 1418 A; 1016 C; 977 G; 1329 T; 0 other;  
 Query Match 14.3%; Score 369.4; DB 23; Length 4740;  
 Best Local Similarity 54.0%; Pred. No. 8.5e-57;  
 Matches 901; Conservative 0; Mismatches 741; Indels 27; Gaps 6;

QY 1151 AGGCTGCTCTTCCAGACACCGCGAGATTCTG9363CCCGCCAGATACAGGCTGTGCT 1210  
 DB 2231 GGGGTGATGTGTGTAGAGATCC---CTTTGACACGACATGCTGAGCTTTCCT 2175  
 QY 1211 GGGAGCTGTGCCCCGGGCGGCTCCGTGGTTCGACCTGTTTGTGAGAGCCAGCTGT 1270  
 DB 2174 CAGGCTGACACAGTGTGTGCAATTGTTGTCTGCAATCTGACAGAGGAACATTTCCGA 2115  
 QY 1271 GTATACCCGACGTCTCTTCCAGGCGCAGCCCTTCATCTGTGTGATGCTGACAACTT 1330  
 DB 2114 ATACAGACTAACCGCTCTCTTATTTGGGACGCCCTTATATGATGCTGATCTTACAACTT 2055  
 QY 1331 TGGGGAAACATGATCTTTTGGAGCCCTAGAGGCTGTGAACGAGGCCGAAACTCTG 1390  
 DB 2054 CGGGGCACTGTGAATGTTTGAATAGCGAAGCTGATTAATCCGGAATAGAAAGGC 1995  
 QY 1391 CCGCTCTTCCCACTCCACATGTGTAGCAGGAGCATGCCCCGAGGGCATCAGCA 1450  
 DB 1994 GCGTCGCTTCCCAACAGCAGTTTGGTGGCAAGGAATCACCCGAGGGCATTTGCGCA 1935  
 QY 1451 GAACGAAGTGTATTTCTTCCCTCATGTGCTAGCTGAGCTGGCGAAAGACCCAGTCCAGA 1510  
 DB 1934 GAATATGTATGTATTCGTTCACTTGAGCGCGGCTGAGATA--ATAATCCCTTAA 1878  
 QY 1511 TTGAGCAGCTGTGATGACAGCTTTGCGCGCGGTATAGGGGTCTCCACCCGAGCG 1570  
 DB 1877 TCTGATAGCTGTATTCGAATTTCTCACAGCCGATATGGCTTAAGATAGGCTTT 1818  
 QY 1571 AGGGGACGCTGAGAGCTACTGCTCCGAGTGTGAACATGCTCCGGGAGGCTTGACG 1630  
 DB 1817 GAGACAGGCTGTGCTGCTCTTAAAGATAGCGCTATTCCTTCGTGCTGCGAGAAAT 1758  
 QY 1631 GGGCGCAATCTGATGAGCCGCTGTGACAGGGGCGGCTGCTACAGATTAATACAGATTCG 1690  
 DB 1757 GCGGGACATATATGATGACGCGCA--GCTTCTTCAACAAAGACCTTTCCTTG 1701  
 QY 1691 GTACAAACCATCTGATGTGTTTGAAGCTGAGCGGCTGTCTCATCTGTCTCCCTCT 1750  
 DB 1700 GTACAAATGCAATGACAGTGTGATGCTGAGCTGTGCTGCTGCTGCTGAGACCATCAT 1641  
 QY 1751 GGC-----CACAGCCCCGCTTCCGCTACAGACCTGTGACCTGACCTGAGCA 1798  
 DB 1640 ACCGCTAGAGATTAACAGATACAGATATACAGAACGATTTGTGTGAATACAGAGCA 1581  
 QY 1799 GAGAGTGCAGAGCTGTGATGCTTGTACTATGAGAGGAGGAGAGGCGCTTACCTGAGCAA 1858  
 DB 1580 ATTCTGAGATCAGGACAGATCAGTTGATATCAATCTAAGGTGGCGTA--TAGAAA 1524  
 QY 1859 GAGAGTGGCTCCCTGTTGAGGAGCTGAGGCGTCTGAGCTATGAGCTGTGCGGCACT 1918  
 DB 1523 GCAACAGGTTTCCGTTTGAAATTCCTTAGCCGCAAGCTGTGAAAATCTGTGATGACAT 1464  
 QY 1919 GAGACGAGTGTGAGTGTGAGCAGCGCTTCTGTGAGGAGGCTGAGAGAGGCGCG 1978  
 DB 1463 GAGCTGATCTGCGGACAGCAGAGAAATTTCTGCTCGGCACTGTGCTCAGAGGCCAA 1404  
 QY 1979 AGAGGCGGAGTGAAGGCGGAGCGCAATTTCTACAGACGAGAACGCCCTTACAGCT 2038  
 DB 1403 GCAAGCGGCGCCCAACATCTGGGCAACAAAGAACTTGATTCATGCGAGAAACCAAT 1344  
 QY 2039 GACCTTGGGGGCGCAGAAAGCAACATCTGTGACATGCAACAGACAGCTGGCGGGTT 2098  
 DB 1343 TACCGCTGGGGTCTGATGTGATGATCTGACATCTGACATATGCTGCAACATGTGCTCGGCTT 1284  
 QY 2099 GGTGGCACTACTACACCCCTGCTGCGGCTTTTCTGTGAGGCGCTG 2147  
 DB 1283 GTTGAACGACTACTATAGGCTGTGAGAGACTTCTCTGGAAGATGTG 1235

RESULT 5  
 AB032798  
 ID AB032798 standard; DNA; 706 BP.  
 XX

AC ABQ32798;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19389.  
 XX  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 XX drug; side effect; cancer; central nervous system; cardiovascular;  
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
 XX SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 EN WO200218632-A2.  
 PD 07-MAR-2002.  
 XX  
 XX 01-SEP-2001; 2001WO-EP10074.  
 XX  
 XX 01-SEP-2000; 2000DE-1043826.  
 XX  
 XX 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 DR WPI; 2002-371829/40.  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX  
 XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's) and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ5411 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 CC  
 CC Sequence 706 BP; 76 A; 93 C; 276 G; 261 T; 0 other;  
 XX  
 SQ  
 Query Match 11.9%; Score 305.2; DB 24; Length 706;  
 Best Local Similarity 77.8%; Pred. No 2e-45;  
 Matches 381; Conservative 0; Mismatches 108; Indels 1; Gaps 1;  
 QY 3 CGGCGTTAGCCTTCGGGTCCACGTGGCGGAGCGCGGACGTGATGAGCGCGGCGCC 62  
 DB 95 CGGCGTTAGTTTTCGGGTTACGTGTCGGA-GTCGGTATGATGATGAGCGCGGTCGTT 153  
 QY 63 CCACCCCTGGCGGTCCGGGACCGGAGACTGAGCCATGAGAGCGGCGGCGGCGC 122  
 DB 154 TTATTTTGTGTCCTCCGGGATTCGATGAGTTGAGTTGAGGCGGCGGCGGTC 213  
 QY 123 GCGCGGTGGGCGCTTCTCTCTGCGCGGCGCGCGCGCGGCGAGCGAGCGCGCG 182  
 DB 214 GCGCGGTGGGCGTTTCTTTTGTGCGGCGTCCGGGCGCGGATGAGCGAGGTTCCG 273

QY 183 GAGCGCGCGCGCGGTGGCGGCGCTGTCGCGCCGCGTGTGCGGCGAGCGCCCGCGAC 242  
 DB 274 GAGCGCGCGGTGTCGCGGCGCTTCGTCGTCGTTGCTGTTGAGTTTCGCGCGAT 333  
 QY 243 TTCTCCGTGTGCGTGGAGCGCGCTCTGCTGCCAACCAGGCGTTGACACTACAGCTTG 302  
 DB 334 TTTTGTGTCGCGTGGAGCGCGCTTTCGTTGTTAGTCGGGTTGATTTATATAGTTTG 393  
 QY 303 GCG 362  
 DB 394 GCG 453  
 QY 363 GCGCTGCACCGGTACCTGCGCGCACTTCTGTCGCTGCCACGTGCGCTCCGCTCTAG 422  
 DB 454 GGGTTATACGTTATTTGCGGATTTTGTGTTGATGCTGTTGTTGTTGTTTNG 513  
 QY 423 CTCGCGCTCCCGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 482  
 DB 514 TTGCGTTGTGCGCGGTTATTTGATGTCGTGCGGCGGATGATGAGGTTACGTTTAT 573  
 QY 483 AGGTACCGCT 492  
 DB 574 AGGTATCGTT 583

RESULT 6  
 ABQ32799/c  
 ID ABQ32799 standard; DNA; 706 BP.  
 XX  
 XX ABQ32799;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19390.  
 XX  
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 XX drug; side effect; cancer; central nervous system; cardiovascular;  
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
 XX SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 EN WO200218632-A2.  
 PD 07-MAR-2002.  
 XX  
 XX 01-SEP-2001; 2001WO-EP10074.  
 XX  
 XX 01-SEP-2000; 2000DE-1043826.  
 XX  
 XX 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 XX  
 DR WPI; 2002-371829/40.  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX  
 XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method



QY 422 GCTGCGCTGCGCGGCGACATGCGCGGCGGAGAGTACCGAGGCCACGCCCA 481  
 DB 607 GTTCCGTTTTCGCGGTTTATTGTTAGTGTGTGCGGCGAGTTATCGAGTTTACGTTTAA 666  
 QY 482 CAGGTACCGCT 492  
 DB 667 TAGGTATCGTT 677

## RESULT 8

ABQ14173/c  
 ID ABQ14173 standard; DNA; 710 BP.

ABQ14173;  
 AC 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 764.

XX Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;  
 XX drug; side effect; cancer; central nervous system; cardiovascular;  
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
 XX SNP; cell differentiation; ds.

OS Homo sapiens.

PN MO200218632-A2.

PD 07-MAR-2002.

PE 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPIDENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CPG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridized to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridization to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridized to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 710 BP; 262 A; 276 C; 91 G; 78 T; 3 other;

XX Query Match 11.8%; Score 304.2; DB 24; Length 710;  
 XX Best Local Similarity 77.4%; Pred. No. 3,1e-45;

Matches 380; Conservative 0; Mismatches 110; Indels 1; Gaps 1;  
 QY 3 CGGGCTTACCTTCCGGTCCACGTGCGCGAGCGCGACGCTGATTTGACGCGGCCGC 62  
 DB 524 CGGGTTTATTTTTCGGGTTTACGTGTCGAGTGGTGGATTTGATGACGCGGGTCTT 465  
 QY 63 CCACCCCTGCGCGCTGCGCGGACCGCAGACTGAGACCATGAGAGCGGTGCGGTC 122  
 DB 464 TTAATTTTGTGTCGTGCGCGGATTTGAGATTGAGATTATGAGAGCGGTGCGGTC 405  
 QY 123 GCGGCGGTGCGGCGCTTCTCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182  
 DB 404 GCGGCGGTGCGGCGCTTCTCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 345  
 QY 183 GAGGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 242  
 DB 344 GAGGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCGCGCGCG 285  
 QY 243 TTCTCGTGTGCGTGGAGCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 302  
 DB 284 TTTTTCGTTGCGTGGAGCGCGCTTGTGTTGTTAGTGGGTTTGGATATTATATTG 225  
 QY 303 GCGGCG 362  
 DB 224 GCGGCG 165  
 QY 363 GGGCTGCAACCGCTACCTGCGCGACTTCTGTGACTGCCAAGCGGCTTGACACTCA 421  
 DB 164 GGGTTGTATCGTTATTTGCGGATTTTGTGTTGTTAGTGGGTTTGGATTTTGA 105  
 QY 422 GCTGCGCGCTGCG 481  
 DB 104 GTTGCCTTGTGCGCGGTTATTGTTAGTGTGTGCGGCGGAGTTGATCGAGTTTAA 45  
 QY 482 CAGGTACCGCT 492  
 DB 44 TAGGTATCGTT 34

## RESULT 9

ABQ32800/c  
 ID ABQ32800 standard; DNA; 706 BP.

ABQ32800;

DE 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19391.

XX Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;  
 XX drug; side effect; cancer; central nervous system; cardiovascular;  
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
 XX SNP; cell differentiation; ds.

OS Homo sapiens.

PN MO200218632-A2.

PD 07-MAR-2002.

PE 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPIDENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of

PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56bp + Sequence Listing; 56bp; German.  
 XX

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 706 BP; 87 A; 93 C; 267 G; 259 T; 0 other;

XX Query Match 10.1%; Score 259.8; DB 24; Length 706;  
 XX Best Local Similarity 71.9%; Pred. No. 2.5e-37;  
 XX Matches 353; Conservative 0; Mismatches 137; Indels 1; Gaps 1;

QY 1 CCGGGGCTTAAAGCTTCCGAGTGGCGGAGGCGCGGACGCTGATGAGCGGGGCG 60  
 DB 614 CCCGAACTTAACCTTCGATCCAGCTAAC- GAAACCGACACTAATTAAACGGAACCG 556  
 QY 61 CCCCAACCCCTGCGCGTGGCGGACCCGAGACTGAGACCATGAGCGGTGGCGTGG 120  
 DB 555 CCCCAACCCCTTACCGTGGCGGAAACCGGAAACCTAATAAACGATTAACGATTA 496  
 QY 121 CCGGGGCGGTGGGGGCTTCTTCCTGCGCGGGGCGGGGGCGCGGACGACGAGCGCC 180  
 DB 495 CCGGACGATTAATAAATCTTCTTACCGGAAACCGAAACGCGCAACGACGAAACCC 436  
 QY 181 GCGAGCGCGCGCGCTGCGGCGCTGCTGCGCGCTGCTGCGCGCGCGCGCGCGCG 240  
 DB 435 GAAAAACGACGACCGGACGACGCTGTAACCGGACTTAACCAAAACCGCGGACCG 376  
 QY 241 ACTTCTCGGTGCGGTGAGCGCGCTGCGTCCGACCGCGGCTTGAACACTTAACCG 300  
 DB 375 ACTTCTCGGTGCGGTGAGCGCGCTGTAACCAAAACCGAACTTAACCTTAACCG 316  
 QY 301 TGGGCGCGCGCGCGCGCGCGCGCTGCGGCGTGGCTCCACGCGGCGTGGCGCGCG 360  
 DB 315 TAAAGGACGACGACCGGACGCGCGCTGTAACCGGACTTCAACGAAACGTAACCGCG 256  
 QY 361 GCGGCGTGAACCGCTTAACCTGCGGACTTGTGCTGCGACGCTGCGCTGCTGCTTC 420  
 DB 255 CGAAACTACACCGCTTAACCTGCGGACTTGTATTAACGCTTAACCTTAACCTGCTTC 196  
 QY 421 AGCTGCGCGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
 DB 195 AACTAGCGCTTACCGGACGACCTTAACCAACGTAACCGAAACCAAACTAACCGAAC 136  
 QY 481 ACAAGTACCGC 491  
 DB 135 ACAAAATACCGC 125

RESULT 10  
 ABQ32801  
 ID ABQ32801 standard; DNA; 706 BP.  
 XX  
 AC ABQ32801;

XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 19392.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 XX drug; side effect; cancer; central nervous system; cardiovascular;  
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
 XX SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Gueitig D;

DR WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful  
 XX for diagnosis and prognosis, comprises selective hybridization of  
 XX amplicons from chemically treated DNA -

XX Claim 12; 56bp + Sequence Listing; 56bp; German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX Sequence 706 BP; 259 A; 267 C; 93 G; 87 T; 0 other;

XX Query Match 10.1%; Score 259.8; DB 24; Length 706;  
 XX Best Local Similarity 71.9%; Pred. No. 2.5e-37;  
 XX Matches 353; Conservative 0; Mismatches 137; Indels 1; Gaps 1;

QY 1 CCGGGGCTTAAAGCTTCCGAGTGGCGGAGGCGCGGACGCTGATGAGCGGGGCG 60  
 DB 93 CCCGAACTTAACCTTCGATCCAGCTAAC- GAAACCGACACTAATTAAACGCGGACCG 151  
 QY 61 CCCCAACCCCTGCGCGTGGCGGACCCGAGACTGAGACCATGAGCGGTGGCGTGG 120  
 DB 152 CCCCAACCCCTTACCGTGGCGGAAACCGGAAACCTAATAAACGATTAACGATTA 211  
 QY 121 CCGGGGCGGTGGGGGCTTCTTCCTGCGCGGGGCGGGGGCGCGGACGACGAGCGCC 180  
 DB 212 CCGGACGATTAATAAATCTTCTTACCGGAAACCGGAAACCGGACGACGAAACCC 271  
 QY 181 GCGAGCGCGCGCGCTGCGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240

Db 272 GAAAGACGAGCGGTAAGGAGCGGTAACCGGACTTAAACCAACCCCGGACCG 331  
 QY 241 ACTTCTCCGTGCGTGTAGAGCGCGCTGTGGCTGCAAGCGGGCTTGGACACCTAGCGC 300  
 Db 332 ACTTCTCCGTATGATTAATAAGCGGCTTAAGTACCAACCGAACTTAAACCTTACAAAC 391  
 QY 301 TGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
 Db 392 TAAACGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 451  
 QY 361 CGGGGCTGACCGCGTACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
 Db 452 CGAAACCTACACCGCTACTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 511  
 QY 421 AGCTGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480  
 Db 512 AACTACCGCTTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 571  
 QY 481 ACAGGTACCGCG 491  
 Db 572 ACAATATCCGC 582

RESULT 11  
 ABQ14170/c  
 ID ABQ14170 standard; DNA; 710 BP.  
 XX

12-JUL-2002 (first entry)  
 XX

Oligonucleotide for detecting cytosine methylation SEQ ID NO 761.  
 XX

KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KM drug; side effect; cancer; central nervous system; cardiovascular;  
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KM SNP; cell differentiation; de.  
 XX

OS Homo sapiens.  
 XX

PN W0200218632-A2.  
 XX

PD 07-MAR-2002.  
 XX

PF 01-SEP-2001; 2001WO-BP10074.  
 XX

PR 01-SEP-2000; 2000DE-1043825.  
 XX

PR 05-SEP-2000; 2000DE-1044543.  
 XX

PA (EPIC-) EPIDEMIOLOGICS AG.  
 XX

PI Olek A, Piepenbrock C, Berlin K, Gnetig D;  
 XX

DR WPI; 2002-371829/40.  
 XX

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX

PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridized to two classes, each with at least one  
 CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridization to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridized to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of

CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 XX

SQ Sequence 710 BP; 103 A; 91 C; 250 G; 263 T; 3 other;

Query Match 10.0%; Score 257.2; DB 24; Length 710;  
 Best Local Similarity 71.3%; Pred. No. 7.3e-37;  
 Matches 351; Conservative 0; Mismatches 140; Indels 1; Gaps 1;

QY 1 CCGGGGCTTACGCTTGGGCTCCAGGTGGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGG 60  
 Db 526 CCGGACCTTAACCTTGCATCCAGTAAACGAAACGACAACTTAATTAACGGAAACCG 467  
 QY 61 CCCCACCCCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120  
 Db 466 CCCCACCCCTTAAACCGTGCAGAAACCGGCAAACTTAAACCATTAATAAGTAAGATAA 407  
 QY 121 CCGCGGCGGTGGGGGCTTCTTCTTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180  
 Db 406 CCGGACGATTAATAATCTTCTTCTTAAACGAAACGAAACGAAACGAAACGAAACCG 347  
 QY 181 GGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
 Db 346 GAAAGACGACCGCTAGGAAACGCTCGTAACCGGACTTAACCTTAATAACCAAAACCGG 287  
 QY 241 ACTTCTCCGTGTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
 Db 286 ACTTCTCCGTGTGAGTAATAACGCGCTTAACTTAACCAAAACGAACTTAATAACCTTAAC 227  
 QY 301 TGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
 Db 226 TAAAGACGAGGAAACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 167  
 QY 361 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 419  
 Db 166 GAAACCTACACCGCTACTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 107  
 QY 420 CAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 479  
 Db 106 CAACTACGCGCTTACCGGCACTACCAACCGTACGAAACCACTTAACCGAAACCGGCGG 47  
 QY 480 AACAGGTACCGC 491  
 Db 46 AACAAATACCGC 35

RESULT 12  
 ABQ14171

ID ABQ14171 standard; DNA; 710 BP.  
 XX

AC ABQ14171;  
 XX

DT 12-JUL-2002 (first entry)  
 XX

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 762.  
 XX

KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KM drug; side effect; cancer; central nervous system; cardiovascular;  
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KM SNP; cell differentiation; de.  
 XX

OS Homo sapiens.  
 XX

PN W0200218632-A2.  
 XX

PD 07-MAR-2002.  
 XX

XX 01-SEP-2001; 2001WO-EP10074.  
 PF 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX (EPiG-) EPIGENOMICS AG.  
 PA Olek A, Piepenbrock C, Berlin K, Guelzig D;  
 PI MPI, 2002-371829/40.  
 DR  
 XX  
 XX  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 CC  
 XX  
 SQ Sequence 710 BP; 263 A; 250 C; 91 G; 103 T; 3 other;  
 Query Match 10.0%; Score 257.2; DB 24; Length 710;  
 Best Local Similarity 71.3%; Pred. No. 7.3e-37;  
 Matches 351; Conservative 0; Mismatches 140; Indels 1; Gaps 1;  
 QY 1 CCGGGCTTACCTTCGGGTCCAGTGGCCGAGCCGAGCTGATGAGCGGCGCG 60  
 DB 185 CCGGAACTTAACCTTCGATCAAGTAACGAAACCGAACCTTAATTAACGGAACCG 244  
 QY 61 CCGCAACCCCTGCGCGGTGCGGAGCCGCGAGACTGAGACCGGTGCGGTG 120  
 DB 245 CCGCAACCCCTTAACCGTGGGAAACCGGAAACCTTAATTAACGATTAACGATTA 304  
 QY 121 CCGGCGCGGTGGGGGTCTTCTTCCTGAGCGGGGGCGGGGCGGAGCGAGCGCC 180  
 DB 305 CCGGACCATATAAATCTTCTCTTAACGAAACCGAAACCGAACGACGAAACCGCC 364  
 QY 181 GCGAGCGCGCGCGCGGTGCGGCGGTGCGGCGCGGTGCGGCGCGCGCGCGCG 240  
 DB 365 GAAACACGACGACGCTAGCAAGCGCTCGTAACCGACCTTAATTAACCGCGCGCG 424  
 QY 241 ACTTCTCCGTGCGGTGAGCGCGCTGAGTGCAGACCGGGCTTGAACACTTAACGCC 300  
 DB 425 ACTTCTCCGTATCGATATAAAGCGCTCTTAACCTTAACCGAACTTAACACTTAAC 484  
 QY 301 TGGGCG 360  
 DB 485 TAAACAGAGAGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 544  
 QY 361 CCGGCG 419  
 DB 545 CGAAACTACACCGCTACTACGCACTTCTAATTAACGATTAACGATTAACGACTCT 604

QY 420 CAGTGGCGCTTGGCGGCGCACTGCGGAGCGGTGCGGGGAGCTGACCGAGCGCC 479  
 DB 605 CACTACGCGCTACCGCGAGCACTACCAACCGGAGGAACTAATTAACCGAGCGCC 664  
 QY 480 AACAGTACCGCG 491  
 DB 665 AACAAATACCGC 676  
 RESULT 13  
 ABL22584/c  
 ID ABL22584 standard; DNA; 7222 BP.  
 XX  
 XX ABL22584;  
 AC  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 19225.  
 XX  
 XX Drosophila melanogaster genomic polynucleotide; cell signalling; insecticide;  
 KW pharmacological; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN MO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PMD, Myers EW,  
 DR MPI, 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS  
 PS Claim 1; SEQ ID NO 19225; 21pp + Sequence Listing; English.  
 CC  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
 CC sequences (AB01840-AB116175) and the encoded proteins  
 CC (AB057737-AB072072).  
 CC  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 7222 BP; 2085 A; 1560 C; 1634 G; 1943 T; 0 other;  
 Query Match 3.5%; Score 89.2; DB 23; Length 7222;  
 Best Local Similarity 54.3%; Pred. No. 7.1e-07;  
 Matches 203; Conservative 0; Mismatches 168; Indels 3; Gaps 1;  
 QY 1774 AGACCTGTGACCTCACTGCGAGGAGTGCAGAGCTGTGACCTGTGACTGAGG 1833  
 DB 7145 ACGATTGGTGAATTAACAGGCAATTCGCAATGATGAGCGAGATGATGATGATCA 7086  
 QY 1834 AGCAAGAAAGCCCTTAACCTGAGCAAGAGCTGCGCTCTGTTGAGGCTGAGAGCGTCC 1893  
 DB 7085 ATCTAAGGTGGCTTA--TGAAGAGCAAGATTGCGCTTTCAATTCCTTAAGCTCA 7025  
 QY 1894 TGGCTATAGAGTGTGCGCGGCACTGAGCAAGGTGCTGCTAGTACAGCGGCTTCTTGC 1953

Db 7028 AGCTGTGAAACTGTTGATGATGAGTGTGCTGGCGAGCAGAGAAATTTCTGC 6369  
 Qy 1954 TGGGACAGTGGCTAGAGAGGCGCCGAGAGCAGTCACTGAGAGCCGAGCCGATTTCT 2013  
 Db 6968 TCGGCACTGGCTCAGAGAGGCGCCGAGAGCAGTCACTGAGAGCCGAGCCGATTTCT 6309  
 Qy 2014 ACGGACAGAGAGCGGCTCAGAGTGTGCTGGGAGCGAGAGCAATCTGACT 2073  
 Db 6908 TCGGATTCATGCGAGAAACCAATTAACCGCTGGGCTGATGTGCAATCTGACT 6849  
 Qy 2074 ATGGCAACAAGCAGTGGCGGAGTGTGAGCACTACACCCCTGGCTGGCGGCTT 2133  
 Db 6848 ATGGCTGCAACATGTGCTCCGCTTGGTGAAGCACTACTATAGGCTTGGTGAAGCTCT 6789  
 Qy 2134 TCTTGAAGCGCTG 2147  
 Db 6788 TCTTGAAGATGTG 6775

RESULT 14  
 AAX53491/C  
 ID AAX53491 standard; DNA; 114955 BP.  
 AC AAX53491;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX  
 DE Human adenosine A1 receptor antisense oligonucleotide fragment.  
 XX  
 KW Antisense oligonucleotide; multiple target; antisense treatment;  
 KW impaired respiration; inflammation; lung disease;  
 KW pulmonary vasoconstriction; inflammation; allergic rhinitis;  
 KW acute asthma; allergy; asthma; impeded respiration;  
 KW respiratory distress syndrome; pain; cystic fibrosis;  
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;  
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;  
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;  
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;  
 KW prostate cancer; ss.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9913886-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 17-SEP-1998; 98WO-US19419.  
 XX  
 PR 09-JUN-1998; 98US-0093972.  
 PR 17-SEP-1997; 97US-0059160.  
 XX  
 PA (UYEC-) UNIV EAST CAROLINA.  
 XX  
 PI Nyce JW;  
 DR WPI; 1999-229400/19.  
 XX  
 DE New antisense oligonucleotides used in treatment of, e.g. pulmonary  
 PT vasoconstriction  
 XX  
 XX Disclosure; Page 37, 120pp; English.  
 XX  
 CC The specification describes antisense oligonucleotides (AAX52869-X55271)  
 CC directed against at least 2 mRNAs selected from target genes, coding and  
 CC non-coding regions of RNAs corresponding to target genes, gene  
 CC initiation codons, genomic flanking regions, intron-exon borders, the  
 CC 5'-end, the 3'-end and the junction between coding and non-coding  
 CC regions and all segments of RNAs encoding proteins associated with one  
 CC or more diseases, conditions or mixtures. The antisense oligonucleotides  
 CC may be derived from sequences AAX55272-74. These multiple target  
 CC oligonucleotides (specifically AAX55180-271) can be used for the  
 CC antisense treatment of diseases and conditions. Typical diseases and  
 CC conditions are those associated with impaired respiration and

CC inflammation, including lung diseases, pulmonary vasoconstriction,  
 CC inflammation, allergic rhinitis, acute asthma, allergies, asthma, impeded  
 CC respiration, respiratory distress syndrome, pain, cystic fibrosis,  
 CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic  
 CC obstructive pulmonary disease (COPD), and cancers such as leukemias,  
 CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,  
 CC pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,  
 CC or have metastasized to the lungs, including breast and prostate cancer.  
 XX

Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;  
 SO

Query Match 3.1%; Score 80.4; DB 20; Length 114955;  
 Best Local Similarity 34.0%; Pred. No. 3.1e-05;  
 Matches 159; Conservative 55; Mismatches 253; Indels 0; Gaps 0;

Qy 25 GTGGCGGAGCGGCGAGCTGATGAGCGGCGCGCCGACCCCTGGCCGCTCGCGGA 84  
 Db 104878 GNNHNNNSVGGCCVCGCGGNNHNNNSVGGCCVCGCGGNNHNNNSVGGCCVCGCGGNN 104819  
 Qy 85 CCCGAGAGCTGAGACATGAGGCGGTGGCGGTGGCGCGCGGTGGCGGTCTTCTCC 144  
 Db 104818 NSCCVGGCCVCGCGGNNHNNNSGCCVGGCCVCGCGGNNHNNNSGCCVGGCCVCGGNN 104759  
 Qy 145 TGGCCGCGCGCGCGCGCGCGCGAGCGAGCGCCCGGAGCGCGCGCGCGCGCGCG 204  
 Db 104758 HNNSGCGCCVCGCGCGGNNHNNNSCGCGCCVCGCGGNNHNNNSGCGCGCC 104699  
 Qy 205 TCGTGGCCCGCGCTGTGTGGCGCGAGCGCCCGCGCGGCTTCTGCTGTGTGAGAGCGG 264  
 Db 104698 VGGCCVCGGNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104639  
 Qy 265 CTCTGCTGTGCAAGCGCGCTTGTGACACTTACAGCTGGCGCGCGCGCGCGCGCG 324  
 Db 104638 CGNNHNNNSGCG 104579  
 Qy 325 TCGGAGTGTGCGCGCTGTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 384  
 Db 104578 NNHNNNSCG 104519  
 Qy 385 ACTTGTGTGCTGTGCAAGCGCGCTGTGTGTGCGCTGTGCGCGCGCGCGCGCGCG 444  
 Db 104518 GNNHNNNSGCGCGCGCGCGCGCGCGCGCGCGCGCGGNNHNNNSGCGCGCGCGCGCGCG 104459  
 Qy 445 CAGCCGCGCGCGGAGCTGACCGAGCGCCAGCGCCCAAGATACCGC 491  
 Db 104458 CVGCGGNNHNNNSCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104412

RESULT 15  
 AAT91361/C  
 ID AAT91361 standard; DNA; 4020 BP.  
 XX  
 AC AAT91361;  
 XX  
 DT 14-APR-1998 (first entry)  
 XX  
 DE Orf virus genomic DNA BamHI B and BamHI G fragments.  
 XX  
 KW Parapoxvirus; vector; vaccine; antigen; HIV; herpes simplex virus;  
 KW Echinosococcus granulosis; Trichostromylinus; Haemophilus;  
 KW Osterlagia; Taenia ovis; ds.  
 XX  
 OS Orf virus strain NZ-2.  
 XX  
 FH Key location/Qualifiers  
 FT CDS complement (27..589)  
 FT /\*tag= a  
 FT /\*label= E1L  
 FT /\*note= "putative E1L coding sequence"  
 FT promoter complement (662..694)  
 FT /\*tag= b  
 FT /\*label= PE1L

FT		/note= "putative E1L gene promoter"
FT	misc_feature	736..1439
FT		/*tag= c
FT		/note= "intergenic region suitable for foreign
FT	CDS	gene insertion"
FT		complement (1440..1661)
FT		/*tag= d
FT	misc_feature	/note= "putative E2L coding region"
FT		1662..1726
FT		/*tag= e
FT		/note= "intergenic region suitable for foreign
FT	promoter	gene insertion"
FT		complement (1708..1732)
FT		/*tag= f
FT		/label= PE2L
FT	CDS	/note= "putative E2L gene promoter"
FT		complement (1726..2224)
FT		/*tag= g
FT		/label= E3L
FT	misc_feature	/note= "putative E3L coding region"
FT		2226..2286
FT		/*tag= h
FT		/note= "intergenic region preferred for foreign
FT	promoter	gene insertion (Claim 13) "
FT		complement (2243..2276)
FT		/*tag= i
FT		/label= P33L
FT		/note= "putative E3L gene promoter"
FT	CDS	complement (2287..3837)
FT		/*tag= j
FT		/label= G1L
FT		/note= "putative G1L coding sequence"
FT	promoter	complement (3838..3876)
FT		/*tag= k
FT		/label= PG3L
FT		/note= "putative G1L gene promoter"
FT	misc_feature	3838..4009
FT		/*tag= l
FT		/note= "intergenic region suitable for foreign
FT		gene insertion"
XX	M09737031-A1.	
BD	09-OCT-1997.	
XX		
PF	27-MAR-1997;	97WO-NZ00040.
XX		
XX	29-MAR-1996;	96NZ-0286284.
PA	(UYOT-) UNIV OTAGO.	
P1	Lytic Df, Robinson AJ;	
DR	WPI; 1997-503116/46.	
XX		
XX		
CC	This genomic DNA molecule comprises the BamH I and BamH I G	
CC	fragments of the orf virus strain NZ-2 genome incorporating	
CC	intergenic regions suitable as foreign gene insertion sites. A	
CC	claimed parapoxvirus vector comprises a parapox virus (especially	
CC	an orf virus) containing exogenous DNA. Vector fragments or	
CC	variants with equivalent immunological activity and host cells	
CC	(especially bovine or ovine testis cells) incorporating the vector	
CC	are also claimed. The vector and fragments/variants are useful in	
CC	vaccine production (claimed), since they can express antigens,	
CC	useful in medical and veterinary applications e.g. parasitic	
CC	disease control in livestock. Particular antigens include HIV	
CC	envelope protein, herpes simplex glycoprotein, antigens of	
CC		

CC Echinococcus granulosus, Trichostrongylus, Haemonchus, Oeseragia  
CC and especially Taenia ovis 45W, 16 kD, 18 kD antigens. The foreign  
CC gene is preferably incorporated in a non-essential region of the  
CC virus genome (see AAT91357-61) under control of a poxvirus promoter,  
CC especially of the virus promoter EL1, FL1 or P3R (see AAT91362-64). The  
CC of virus vectors limit the cross-species infection common with e.g.  
CC vaccinia virus vectors, since the virus has a narrow host range;  
CC orf is also less virulent than vaccinia in man.  
CC  
CC  
CC Sequence 4020 BP; 755 A, 1310 C, 1278 G, 677 T, 0 other;  
CC  
CC

Query Match 2.8%; Score 71; DB 18; Length 4020;

Matches 178; Conservative 0; Mismatches 150; Indels 7; Gaps 1;

111 GTGGCGGTGGCCGCGCGGTGGGGTCTTCTCTGCGCCGCGGCGGGGCGCGGCAAGC 170

[illegible]

171 AAAAAAAAAAAAAAAAAAAAAAATTTTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCCC

[illegible][illegible]

231 CCGCGGCGACTTCTCCGTGTGGTGAGCCGCTCTGGCTGCCAAGCCGGCTTGAC 290

b  
922 GGCCCGCTGGACGCGCGCTGCCGCTGCTGCGCGCGCGCGCGCGCGCGCGCTCGCG 863

Y  
291 ACCTACAGCCTGGGCGGCGGCGGCGGCGGCGGTGCCGCTCCACGGGCGTG 350

b  
862 CCGCCGCGCAGCGCGCTCGCGCGCGCGCTGCCTGCCCG 810

351 GCGGCCGCCGGGCTGCACCGCTACCTGCGGACTTCTGTGGCTGCCACGTGGCCTGG 410

809 CCGCGGTGGCCGCGCTGGCCGCGCGCTTGGCCGCCCGCGCGCGCGGTAGCG 750

TCCCCCTCCAGACCTGCGGCCGTGGCGGCAGCATTGCC

441

[illegible][illegible]

Search completed: February 13, 2004, 09:26:48

job time : 682 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 13, 2004, 09:06:34 ; Search time 5297 Seconds

(without alignments)  
11814.995 Million cell updates/sec

Title: US-09-836-613-1

Perfect score: 2575  
Sequence: 1 CCCGGGCTAGCCTTCGGGT.....AGTCAGCGCGCCGCAATTC 2575

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inu:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rnd:\*  
26: em\_gss\_rnd:\*  
27: em\_gss\_vit:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	929.4	36.1	952	12	BM473725	BM473725 AGENCOURT
2	800.2	31.1	904	13	BU182537	BU182537 AGENCOURT
3	791.8	30.7	808	13	BI67002	BI67002 603291867
4	789.6	30.7	925	13	BQ893829	BQ893829 AGENCOURT

5	753	29.2	989	12	BG760547	BG760547 602716985
6	748.6	29.1	770	12	BQ001701	BQ001701 UI-H-DH1-
7	748.4	29.1	757	12	BQ014565	BQ014565 UI-H-BD1-
8	745	28.9	773	12	BM976832	BM976832 UI-CF-EN1
9	743	28.9	788	9	AT118592	AT118592 AT118592
10	741.8	28.8	1097	10	BG327357	BG327357 602426586
11	739.4	28.7	758	12	BM976494	BM976494 UI-CF-EN1
12	722.4	28.1	829	12	BG910492	BG910492 602806176
13	707.6	27.5	816	10	BG476071	BG476071 602520954
14	702	27.3	707	13	BQ772664	BQ772664 UI-H-FE1-
15	701.4	27.2	707	12	BQ006037	BQ006037 UI-H-BD1-
16	682.2	26.6	725	12	BM726804	BM726804 UI-B-BD1-
17	682.4	26.5	707	12	BQ018688	BQ018688 UI-H-FE1-
18	675	26.2	699	13	BQ632129	BQ632129 UI-H-BD1-
19	670.4	26.0	676	13	BM683837	BM683837 UI-CF-ENO
20	663.8	25.4	1201	13	BX417673	BX417673 BX417673
21	655.2	25.4	800	10	CA439269	CA439269 UI-H-DT1-
22	652.6	25.3	800	10	BF338953	BF338953 602036023
23	652	25.3	978	13	BU191481	BU191481 AGENCOURT
24	651.4	25.3	929	10	BG107971	BG107971 602278267
25	649.4	25.2	848	10	BG117672	BG117672 602349632
26	645.4	25.1	692	9	AI436192	AI436192 UI-H-FLO-
27	643	25.0	648	13	BU630517	BU630517 UI-H-FLO-
28	640.6	24.9	974	13	BQ677244	BQ677244 AGENCOURT
29	638.8	24.8	947	13	BU542714	BU542714 AGENCOURT
30	633.4	24.6	741	10	BE410798	BE410798 601301429
31	633	24.6	633	12	BM716556	BM716556 UI-B-DX1-
32	629	24.4	640	10	BF346260	BF346260 602018449
33	627.2	24.4	654	12	BM728086	BM728086 UI-B-EJO-
34	618.8	24.0	655	9	AM170220	AM170220 xm57a12.x
35	614	23.8	625	10	BG149216	BG149216 nad24507
36	613.2	23.8	926	13	BU591032	BU591032 AGENCOURT
37	612	23.8	612	9	AI686304	AI686304 FE33C07.x
38	610.4	23.7	612	9	AM409366	AM409366 FH03G08.Y
39	606.2	23.5	623	12	BM784286	BM784286 K-EST0062
40	596.2	23.2	626	10	BE219892	BE219892 hv64D03.x
41	583	22.6	583	12	BM773017	BM773017 K-EST0057
42	581.2	22.6	824	12	BI557975	BI557975 603237595
43	578.4	22.5	907	13	BX454545	BX454545 BX454545
44	576	22.4	696	14	CB473703	CB473703 SH74.F05
45	570	22.1	956	13	BQ922969	BQ922969 AGENCOURT

## ALIGNMENTS

RESULT 1  
BM473725  
LOCUS  
DEFINITION  
AGENCOURT 6484269 NIH\_MGC\_72 Homo sapiens cDNA IMAGE:55318154  
5', mRNA sequence.  
ACCESSION  
BM473725  
VERSION  
BM473725.1 GI:18522767  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 952)  
NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS  
National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE  
JOURNAL  
COMMENT  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: ATCC/DC/DMP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
plate: LHAM12230 row: j column: 11  
High quality sequence stop: 679.

# FEATURES

Location/Qualifiers  
1. .952  
/organism="Homo sapiens"  
/mol\_type="rRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5538154"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 2 kb. Library constructed by Life  
Technologies"

BASE COUNT 172 a 304 c 288 g 187 t 1 others

ORIGIN

Query Match 36.1%; Score 929.4; DB 12; Length 952;  
Best Local Similarity 99.3%; Pred. No. 2.2e-164;  
Matches 933; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

950 GGAAGACCCCATATTCCTCCATCATGAGAGCCTCTCTCTGAGAGCTGATCAAGAGTT 1009  
13 GGAAGACCCCATATTCCTCCATCATGAGAGCCTCTCTCTGAGAGCTGATCAAGAGTT 72  
1010 TGGCAGACACATCTATGAGGAGCGAGCACTTTCATGAGATGAGCAGCAGCTTCTCAGA 1069  
73 TGGCAGACACATCTATGAGGAGCGAGCACTTTCATGAGATGAGCAGCAGCTTCTCAGA 132  
1070 GCCCTCTACCTTGGCGGAGCCAGCCAGCTGCTATGAGGAGCAGTACTGATGATAC 1129  
133 GCCCTCTACCTTGGCGGAGCCAGCCAGCTGCTATGAGGAGCAGTACTGATGATAC 192  
1130 TGAAGCTGTGTGAGCTGTCTCAAGAGCTGCTCTTCAGACAGCGAGAGTTCTGGGGCC 1189  
193 TGAAGCTGTGTGAGCTGTCTCAAGAGCTGCTCTTCAGACAGCGAGAGTTCTGGGGCC 252  
1190 CGCCCAATCAAGAGCTGTGTGAGAGCTGTGCGCCGCTGCTGCTGATCTGAACT 1249  
253 CGCCCAATCAAGAGCTGTGTGAGAGCTGTGCGCCGCTGCTGCTGATCTGAACT 312  
1250 GTTGTGAGAGCGAGCTGTGTATACCGGAGAGCTGCTCTTCAGAGGCGAGCCCTCAT 1309  
313 GTTGTGAGAGCGAGCTGTGTATACCGGAGAGCTGCTCTTCAGAGGCGAGCCCTCAT 372  
1310 CTGTGATGCTGCAAACTTTGGGGAAACATGCTTTTGGAGAGCCCTAAGAGCTGT 1369  
373 CTGTGATGCTGCAAACTTTGGGGAAACATGCTTTTGGAGAGCCCTAAGAGCTGT 432  
1370 GAAAGGAGGCGGAGAGCTGCGCCGCTCTTCCCAACTCCACATGATGAGGAGGAGCT 1429  
433 GAAAGGAGGCGGAGAGCTGCGCCGCTCTTCCCAACTCCACATGATGAGGAGGAGCT 492  
1430 GAGCCCGAGGAGCATCAGCCAGAGCAAGAGTGTATTCCTCATGCTGAGAGCTG 1489  
493 GAGCCCGAGGAGCATCAGCCAGAGCAAGAGTGTATTCCTCATGCTGAGAGCTG 552  
1490 GGAAGAGACCCAGTGTCCAGATTTGGGAGAGCTGAGTGAACAGCTTTGCGCCGCGGTA 1549  
553 GGAAGAGACCCAGTGTCCAGATTTGGGAGAGCTGAGTGAACAGCTTTGCGCCGCGGTA 612  
1550 TGGGAGTCTCCACCGGAGCGAGGAGGAGCTGAGGAGCTACTGCTCCGAGTGTGTA 1609  
613 TGGGAGTCTCCACCGGAGCGAGGAGGAGCTGAGGAGCTACTGCTCCGAGTGTGTA 672  
1610 CTGCTCCGAGGAGGCTGAGGAGGAGCGCAATGTATGAGGAGGAGGAGGAGCTGCT 1669  
673 CTGCTCCGAGGAGGCTGAGGAGGAGCGCAATGTATGAGGAGGAGGAGGAGCTGCT 732  
1670 ACAGATGAATACAGAGCTGTGTAACAAGATCTGATGTGTTGAGGAGGAGGAGCTGCT 1729  
733 ACAGATGAATACAGAGCTGTGTAACAAGATCTGATGTGTTGAGGAGGAGGAGCTGCT 792  
1730 GCTCAATCTGCTCCCTCCCTGAGCAACAGCCCGGCTTCCGCTAAGAGCTGAGAGCT 1789

Db 793 GCTCAATCTGCTCCCTCCCTGAGCAGAGCCCGGCTTCCCTGAGAGCTGAGCT 852

Qy 1790 CACTGGCAGGAGAGTGCAGAGAGCTGATGAGTGTGATGATGAGAGGAGAGGAGCTTA 1849

Db 853 CACTGGCAGGAGAGTGCAGAGAGCTGATGAGTGTGATGATGAGAGGAGAGGAGCTTA 912

Qy 1850 CCTGAGCAGAGAGTGGCTCCCTGTTGAGGAGCTGAGGC 1899

Db 913 CCTGAGCAGAGAGTGGCTCCCTGTTGAGGAGCTGAGGC 952

RESULT 2

BUI82537 904 bp mRNA linear EST 04-SEP-2002

LOCUS

DEFINITION AGENCOURT\_7946518 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6151074

ACCESSION BUI82537

VERSION BUI82537.1 GI:22696521

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC http://mgi.nci.nih.gov/.

TITLE 1 (bases 1 to 904)

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished

Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: ATCC/DCTD/DTF  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at:  
http://image.llnl.gov  
Plate: LLM13486 row: 1 column: 19  
High quality sequence stop: 598.

Location/Qualifiers

1. .904  
/organism="Homo sapiens"  
/mol\_type="rRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6151074"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 2 kb. Library constructed by Life  
Technologies"

BASE COUNT 171 a 279 c 283 g 171 t

ORIGIN

Query Match 31.1%; Score 800.2; DB 13; Length 904;  
Best Local Similarity 96.1%; Pred. No. 3.9e-140;  
Matches 864; Conservative 0; Mismatches 28; Indels 7; Gaps 4;

Qy 1463 CTATTCCTCATGCTGAGCTGGCTGGCGAAAGACCCAGTGCAGATTTGGAGCTG 1522

Db 1 CTATTCCTCATGCTGAGCTGGCTGGCGAAAGACCCAGTGCAGATTTGGAGCTG 60

Qy 1523 GGTGACACATTTGCGCCGCGGAGATGAGGATCTCCACCGGAGCGAGGAGGAGCTG 1582

Db 61 GGTGACACATTTGCGCCGCGGAGATGAGGATCTCCACCGGAGCGAGGAGGAGCTG 120

Qy 1583 GAGCTACTGCTCCGAGTGTGTAACAATGCTCCGAGGAGGCTGAGAGGAGCAATG 1642

Db 121 GAGCTACTGCTCCGAGTGTGTAACAATGCTCCGAGGAGGCTGAGAGGAGCAATG 180

Qy 1643 TAGCCGCTGTGAGGAGGAGCTCCCTAAGATGAATACAGATCTGTGTAACAAGATC 1702

Db 181 TAGCCCTGCTGAGGCGGCGCTCCCTACAGATGATACAGCATCTGTGTAACAACCGATC 240  
 QY 1703 TGATGTTTGAAGGCGCTGCGGCTGCTGTCAATCTGTCTCCCTCCGCGCAACGAGCC 1762  
 Db 241 TGATGTTTGAAGGCGCTGCGGCTGCTGTCAATCTGTCTCCCTCCGCGCAACGAGCC 300  
 QY 1763 CGCTTCCGCTACACACTGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 1822  
 Db 301 CGCTTCCGCTACACACTGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 360  
 QY 1823 GTACTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1882  
 Db 361 GTACTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420  
 QY 1883 TGAAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1942  
 Db 421 TGAAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 QY 1943 CGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2002  
 Db 481 CGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540  
 QY 2003 GCGCGATTTCTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2062  
 Db 541 GCGCGATTTCTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600  
 QY 2063 CATCTTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2122  
 Db 601 CATCTTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
 QY 2123 CTGCGGCTTTTCTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2182  
 Db 661 CTGCGGCTTTTCTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
 QY 2183 GCACCAAG-TTTGACAAAATGCTTCTTCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2241  
 Db 721 GCACCAAGTTTGAACAAAATGCTTCTTCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780  
 QY 2242 GGTATCCCGAGCGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2300  
 Db 781 GGTATCCCGAGCGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840  
 QY 2301 TATTA--CCCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2354  
 Db 841 TATTAACCCCGAGCGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 899

RESULT 3  
 B1667002 808 bp mRNA linear EST 12-SEP-2001  
 LOCUS 603291867F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:5311121 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B1667002  
 VERSION B1667002.1 GI:15581235  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 808)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strauberg, Ph.D.  
 Email: csabds-remail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki, and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LIML at:  
 http://image.llnl.gov

Plate: LIM11788 row: f column: 18  
 High quality sequence stop: 807.  
 Location/Qualifiers  
 1..808  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5311121"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"  
 /clone\_id="NIH\_MGC\_96"  
 /note="Organ: brain; Vector: pBluescript (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (Gcgag ) ; Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."  
 BASE COUNT 106 a 266 c 303 g 133 t  
 ORIGIN  
 Query Match 30.7%; Score 791.8; DB 12; Length 808;  
 Best Local Similarity 99.6%; Pred. No. 1.5e-138;  
 Matches 804; Conservative 0; Mismatches 2; Indels 1; Gaps 1.

Db 87 CGAGAGCTGAGACCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 146  
 Db 3 CGAGAGCTGAGACCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 62  
 QY 147 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 206  
 Db 63 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 122  
 QY 207 GTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 266  
 Db 123 GTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 182  
 QY 267 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 326  
 Db 183 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 242  
 QY 327 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 386  
 Db 243 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 302  
 QY 387 TTCTGTGCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 446  
 Db 303 TTCTGTGCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 362  
 QY 447 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 506  
 Db 363 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 422  
 QY 507 TGACAGCAAGCTACTCTTCTGT 566  
 Db 423 TGACAGCAAGCTACTCTTCTGT 482  
 QY 567 TGAGTGGGCGGTGAGTGGATGAGTGGATGAGTGGATGAGTGGATGAGTGGATGAGTGGATGAGT 626  
 Db 483 TGAGTGGGCGGTGAGTGGATGAGTGGATGAGTGGATGAGTGGATGAGTGGATGAGTGGATGAGT 542  
 QY 627 CAGCGGAGTGAAGTGGCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 686  
 Db 543 CAGCGGAGTGAAGTGGCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 602  
 QY 687 CCGTGTCTCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 746  
 Db 603 CCGTGTCTCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 661  
 QY 747 CCGTGTCTCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 806

Db	562	CCCTCCTGGGACATCAAGAGAGCTTAACTGACAGACCGGGTCCTGGACCAAGATGCGCTCC	721
Qy	807	TTGGGATATACCCCAAGTGTGCTGCTCATTTGCGGGGGCATGTTCCCGAGGCTGTACCAAG	866
Db	722	TTGGGATATACCCCAAGTGTGCTGCTCATTTGCGGGGGCATGTTCCCGAGGCTGTACCAAG	781
Qy	867	GTGTTCCCTCAGGTCAATGTCAAGAG	893
Db	782	GTGTTCCCTCAGGTCAATGTCCCAAG	808

RESULT	4
BQ893829	
LOCUS	
DEFINITION	BQ893829 925 bp mRNA linear EST 16-AUG-2007 ACENCCOURT_8728368 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:6339115 5' , mRNA sequence.
ACCESSION	BQ893829
VERSION	BQ893829.1 GI:22285843
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo Sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheraia; Primates; Carnivornii; Homnidae; Homo. 1 (bases 1 to 925) NIH-MGC http://mgs.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished
JOURNAL	Contact: Robert Strausberg, Ph.D.
COMMENT	

```

FEATURES
Source
Location/Qualifiers
1..925
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_type="IMAGS:6339115"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="NIH MGC 47"
/note="Organ: brain; Vector: pOTS7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAC(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."
BASE COUNT
164 a 290 c 292 g 179 t

```

	Query Match	30.7%	Score 769.6	DB 13	Length 925
	Best Local Similarity	93.0%	Pred. No. 3.8e-138		
	Matches 893	Conservative	0	Mismatches 9	Indels 58
				Gaps	4
Qy	1114	TCACGTGCGATGATCTAGAGGCTGTGGCTGCTCCAAAGCTGGGCTCTTCCAGACACAGC	1173		
Db	13	TGTCAACGATGATACGTAGGCTGTGGCTGCTCCAAAGCTGGCTCTTCCAGACACAGC	72		
Qy	1174	CGCAGTTCGGGGGGGGCGCCGACAGATCAAGGCTGTGCGGGAGGTGTGCCCCCGGTGGCCCGCC	1233		
Db	73	CGCAGTTCGGGGGGGGCGCCGACAGATCAAGGCTGTGCGGGAGGTGTGCCCCCGGTGGCCCGCC	132		
Qy	1234	TCTCTGTTCTGGACCTGTTTCTAGAGCAGCCTGTGTATACCGCAGCTGGCTCTTTC	1253		

Dp	133	TCCTGATCTCTGAACTGTTTGTGTAGAGCACGCGTGTATACCCGACCTGCTCTTCC	192
Qy	1294	AGGCGACACCCCTTCACTCTGTGCATGCTGCAACACTTTGGGGGAAACATGGTCTTTTGG	135
Dp	193	AGGGCACACCCCTTATCTGTGTGATGTCTGCACAACTTTGGGGGAAACCATGATGCTTTTGG	252
Qy	1354	GAGCCCTTAGAGGCTGTGAACGGAAGGCCCAAGAGTGCCTGGCTCTTCCCAACTCCACCA	1413
Dp	253	GAGCCCTTAGAGGCTGTGAACGGAAGGCCCAAGAGTGCCTGGCTCTTCCCAACTCCACCA	312
Qy	1414	TGTTAGGCAAGGAGCATGGCCCCCGAGGGCATCAGCCAGACGAAGAGTGTCTTATTCCTCA	1473
Dp	313	TGTTAGGCAAGGAGCATGGCCCCCGAGGGCATCAGCCAGAGAGAGTGTCTTATTCCTCA	372
Qy	1414	TGGCTGAGACTGGGCTGGCGAAAGACCCAGTGCAGATTTGGCGAGCTGGGTGACCACT	1533
Dp	373	TGGCTGAGACTGGGCTGGCGAAAGACCCAGTGCAGATTTGGCGAGCTGGGTGACCACT	432
Qy	1534	TTGCCGACCGGCGGATAGGGGCTCTCCACCCCGAGCGACAGGGGCGACGTGAGGGCTACGCG	1593
Dp	433	TTGCCGACCGGCGGATAGGGGCTCTCCACCCCGAGCGACAGGGGCGACGTGAGGGCTACGCG	492
Qy	1594	TCCGAGTGTGTACCACTGCTCCGCGGAGGCGCTGCAAGGCGCAACATGTAGCCCGCTGG	1653
Dp	493	TCCGAGTGTGTACCACTGCTCCGCGGAGGCGCTGCAAGGCGCAACATGTAGCCCGCTGG	552
Qy	1654	TCAGGCGGCGCTCCCTTACAGATGAATATACAGCATCTGTGTACACCCGATCTATGTGTTG	1713
Dp	553	TCAGGCGGCGCTCCCTTACAGATGAATATACAGCATCTGTGTACACCCGATCTATGTGTTG	612
Qy	1714	AGGCTTGCGGCTCTGTGCTCACTCTGTCCCTCCCTGCGACACAGCCCGCGCTTCGCGCT	1773
Dp	613	AGGCTTGCGGCTCTGTGCTCACTCTGTCCCTCCCTGCGACACAGCCCGCGCTTCGCGCT	672
Qy	1774	ACGACTTGCTGGA-CTTCACTTGGCAGGCAAGTGCAGAGACTGTCAAGCTTGTACTATGAG	1832
Dp	673	ACGACTTGCTGGA-CTTCACTTGGCAGGCAAGTGCAGAGACTGTCAAGCTTGTACTATGAG	732
Qy	1833	GAGGCAAGAAAGCGCTTACTGTAGAGGAAGAAGTGTGCTCCCTGTGTGAAGGCTGGAAGCGTC	1893
Dp	733	GAGGCAAGAAAGCGCTTACTGTAGAGGAAGAAGTGTGCTCCCTGTGTGAAGGCTGGAAGCGTC	792
Qy	1893	CTGGCTTATGAGCTGTGCTGCGGCACTGGAAGAGGTGCTGGTATGTGACAGCCGCTTCTTG	1952
Dp	793	CTGGCTTATGAGCT------	806
Qy	1953	CTGGGCAAGCTGGCTTAGAGCAGGCCCGAGCAGCGGCACTCAATGAGG-CCGAGGCCGATTT	2011
Dp	807	-----TGCCTAGAGCAGGCCCGAGCAACGCAATCAGTAGAGCCCGAGGCCGATTT	857
Qy	2012	CTACGAGCAGAAACGCGGCTAACAGCTGACCTTGT-GGGGGCGGAAGGCAACATCTCTGG	2070
Dp	858	CTACGAGCAGAAACGCGGCTAACAGCTGAACTTGTGGGGGCGCGAAGGCAACATCTCTGG	917

RESULT 5	Bg760547	989 bp	mRNA	linear	EST 15-MAY-2001
LOCUS	Bg760547				
DEFINITION	Bg760547	6027216985F1 NIH-MGC_49 Homo sapiens cDNA clone IMAGE:4480456 5',			
ACCESSION	Bg760547				
VERSION	Bg760547.1	GI:14071200			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 989)				
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a>				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgabs-r@mail.nih.gov">cgabs-r@mail.nih.gov</a>				





```

Db      517 TTGTGGGGCCAGAGGCAACATCTGACATATGCAACAGACAGCTGGCGGGTTGTG 458
Qy      2103 GCCACTACTACACACCCCTCGTGCGGCTTTTCTGAGAGGGCTGTGACAGTGGGCC 2162
Db      457 GCCACTACTACACACCCCTCGTGCGGCTTTTCTGAGAGGGCTGTGACAGTGGGCC 398
Qy      2163 CAGGAGCATCCCTTTCCAAACAGACACAGTTTGACAAAATATGCTTCCAACTGAGCAGGCC 2222
Db      397 CAGGAGCATCCCTTTCCAAACAGACACAGTTTGACAAAATATGCTTCCAACTGAGCAGGCC 338
Qy      2223 TTGCTTCTCAGACAGAGGTAACCCAGACAGCGCGGAGAGACACTGTGAGCTGGGCC 2282
Db      337 TTGCTTCTCAGACAGAGGTAACCCAGACAGCGCGGAGAGACACTGTGAGCTGGGCC 278
Qy      2283 AAGAAGATCTTCTCAATATATTACCCGCGCTGGCTGGCCGCTTGTGTATGATTGCG 2342
Db      277 AAGAAGATCTTCTCAATATATTACCCGCGCTGGCTGGCCGCTTGTGTATGATTGCG 218
Qy      2343 CACCACTGGGCGCTTTTCCGCTAATTCCAGGCGCAATTCAGGGCCCGCAGAGCTGGACA 2402
Db      217 CACCACTGGGCGCTTTTCCGCTAATTCCAGGCGCAATTCAGGGCCCGCAGAGCTGGACA 158
Qy      2403 GACATCACAGGATACCCAGGCTGGGAGAGGCCCGCAGCGCTGTGTGGGGCTGGAC 2462
Db      157 GACATCACAGGATACCCAGGCTGGGAGAGGCCCGCAGCGCTGTGTGGGGCTGGAC 98
Qy      2463 CTGGGGGGGATTGAGAGGAATGACCTGCGCTCCACCAACACCAAGTGTGGGATTAAAG 2522
Db      97 CTGGGGGGGATTGAGAGGAATGACCTGCGCTCCACCAACACCAAGTGTGGGATTAAAG 38
Qy      2523 TACTGTTTCTTCCACTTAAAAAATAAAAA 2556
Db      37 TACTGTTTCTTCCACTTAAAAAATAAAAA 4

RESULT 8
BM976832/c
LOCUS      BM976832      773 bp      mRNA      linear      EST 21-FEB-2003
DEFINITION UI-CF-EN1-adt-o-24-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
            UI-CF-EN1-adt-o-24-0-UI.3', mRNA sequence.
ACCESSION  BM976832
VERSION     BM976832.1 GI:19594633
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 773)
            Bonaldo,M.F., Lennon,G. and Soares,M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
            Genome Res. 6 (9), 791-806 (1996)
JOURNAL    97044477
MEDLINE    8889548
COMMENT    Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.reagen.com) or from Open Biosystems
            (www.openbiosystems.com).
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES
SOURCE     location/Qualifiers
            1..773
            /organism="Homo sapiens"

```

```

/mot_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EN1-adt-o-24-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_host="Adult"
/lab_host="DHI0B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EN1"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-EN1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Cystic Fibrosis
Epithelial Cells. The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an EcoR I adaptor, digested with Not
I, and cloned directionally into pT73-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CTGCTCAGGT.
TAG LIB=UI-CF-EN1
TAG_Tissue=Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG_SEQ=CTGCTCAGGT"

BASE COUNT      147 a      232 c      215 g      174 t      5 others
ORIGIN
Query Match      28.9%; Score 745; DB 12; Length 773;
Best Local Similarity 98.6%; Pred. No. 9.1e-130;
Matches 759; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

Qy      1788 CTCATCGGACGACGTGACAGAGCTGTGACCTTGTATGAGAGGCAAGAGCGC 1847
Db      773 CTCATCGGACGACGACGAGAGCTGTGACCTTGTATGAGAGGCAAGAGCGC 714
Qy      1848 TACCTGAGAGAGAGCTGCGCTCCCTGTTGAGGGCTGAGAGGCTGCTGATAGAGCTG 1907
Db      713 TACCTGAGAGAGAGCTGCGCTCCCTGTTGAGGGCTGAGAGGCTGCTGATAGAGCTG 654
Qy      1908 CTGCCGGCACTGAGAGAGTGTGCTAGTAGACACCGCTTCTGCTGGGAGCTGGCTA 1967
Db      653 CTGCCGGCACTGAGAGAGTGTGCTAGTAGACACCGCTTCTGCTGGGAGCTGGCTA 594
Qy      1968 GAGCAGGCGCGGACGCGGCACTGAGAGCGGCGGATTTCTAGAGAGAAAGC 2027
Db      593 GAGCAGGCGCGGACGCGGCACTGAGAGCGGCGGATTTCTAGAGAGAAAGC 534
Qy      2028 CGCTACAGCTGACCTTGTGGGGGCGAGAGGCAATCTGATAGTCCAAAGAGAG 2087
Db      533 CGCTACAGCTGACCTTGTGGGGGCGAGAGGCAATCTGATAGTCCAAAGAGAG 474
Qy      2088 CTGGCGGGGTTGTGGCCAACTACTACACCCCTGCTGGCGGCTTTCTGAGAGCGCTG 2147
Db      473 CTGGCGGGGTTGTGGCCAACTACTACACCCCTGCTGGCGGCTTTCTGAGAGCGCTG 414
Qy      2148 GTTGACAGGTGGGCCCAAGGGATCCCTTCCAAAGACACAGTTTGACAAAATATGCTTC 2207
Db      413 GTTGACAGGTGGGCCCAAGGGATCCCTTCCAAAGACACAGTTTGACAAAATATGCTTC 354
Qy      2208 CAATGAGACAGGCTTGTGTCAGACAGAGGTACCCAGCCAGCCGCGAGAGAGC 2267
Db      353 CAATGAGACAGGCTTGTGTCAGACAGAGGTACCCAGCCAGCCGCGAGAGAGC 294
Qy      2268 ACTGTGAGCTGGGCGCAAGAAATCTTCTCAAAATTTCACCGGCTGGGTGGCGGCTCT 2327
Db      293 ACTGTGAGCTGGGCGCAAGAAATCTTCTCAAAATTTCACCGGCTGGGTGGCGGCTCT 234
Qy      2328 TGGTATGATTCGACACACTGGGCTTGTTCGCTAATTCAGGCGAGATTCCAGG 2387
Db      233 TGGTATGATTCGACACACTGGGCTTGTTCGCTAATTCAGGCGAGATTCCAGG 174

```

QY 2388 GCCCAGAGCTGACAGACATCACAGGATPACCAGCCTCGAGAGAGGCCCCACGGCCTG 2447

Db 173 GCCCAGAGCTGACAGACATCACAGGATPACCAGCCTCGAGAGAGGCCCCACGGCCTG 114

QY 2448 CTGGTGGGGCTCGACCTGGGGGGGATTGGAGGGGAATGACTGGCCCTCCAGCACCACCC-A 2506

Db 113 CTGGTGGGGCTCGACCTGGGGGGGATTGGAGGGGAATGACTGGCCCTCCAGCACCACCCAA 54

QY 2507 AAGTGGGATTAAAGTACTGTTTCTTCCACTTAAAAA 2556

Db 53 AAGTGGGATTAAAGTACTGTTTCTTCCACTTAAAAA 4

RESULT 9  
AUI18592 788 bp mRNA linear EST 01-AUG-2002  
LOCUS AUI18592 HEMBA1 Homo sapiens cDNA clone HEMBA1003940 5', mRNA  
DEFINITION sequence.  
ACCESSION AUI18592  
VERSION AUI18592  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isoagai, T.  
TITLE HRI human cDNA project  
JOURNAL Unpublished  
COMMENT Contact: Takao Isoagai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yama, Kasarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5' - & 3' - end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute  
Location/Qualifiers  
FEATURES  
source 1..788  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HEMBA1003940"  
/cbase\_type="whole embryo, mainly head"  
/dev\_stage="embryo, 10 weeks"  
/clone\_lib="HEMBA1"  
/note="Vector: PME18SFL3"  
BASE COUNT 135 a 250 c 236 g 164 t 3 others  
ORIGIN  
Query Match 28.9%; Score 743; DB 9; Length 788;  
Best Local Similarity 98.3%; Pred. No. 2.2e-129; Indels 1; Gaps 1;  
Matches 760; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

Db 181 CAGATCAGGGCTGTGCTGGAGCTGTGCCCCGTGGCCGCTCCTGATCTGACCTGTT 240

QY 1254 GCTGAGAGCCAGCCTGTGTATACCCGACTGCTCTTTCAGAGGCCAGCCTTCACTGCG 1313

Db 241 GCTGAGAGCCAGCCTGTGTATACCCGACTGCTCTTTCAGAGGCCAGCCTTCACTGCG 300

QY 1314 TGCATGCTGACAACTTGGGGGGAACCATGGCTTTTGAAGCCCTAGAGGCTGTGAAC 1373

Db 301 TGCATGCTGACAACTTGGGGGGAACCATGGCTTTTGAAGCCCTAGAGGCTGTGAAC 360

QY 1374 GAGAGCCAGAGAGCTGCCGCTCTTCCCACTCCACATGTGTAGGACAGGAGCTAGGC 1433

Db 361 GAGAGCCAGAGAGCTGCCGCTCTTCCCACTCCACATGTGTAGGACAGGAGCTAGGC 420

QY 1434 CCCGAGGGCATCAGCCAGAGAGAGTGTATTTCCATGAGCTAGCTGGGCTGGCA 1493

Db 421 CCCGAGGGCATCAGCCAGAGAGAGTGTATTTCCATGAGCTAGCTGGGCTGGCA 480

QY 1494 AAGGACCACTGTCAGATTGGAGGCTGGGTACAGCTTTGCGCCCGCGATAGG 1553

Db 481 AAGGACCACTGTCAGATTGGAGGCTGGGTACAGCTTTGCGCCCGCGATAGG 540

QY 1554 GTCTCCACCCCGAGCCAGAGGAGGAGGCTGAGGCTACTGTCGGAGTGTACAATG 1613

Db 541 GTCTCCACCCCGAGCCAGAGGAGGAGGCTGAGGCTACTGTCGGAGTGTACAATG 600

QY 1614 TCCGGAGAGCTGTGAGGGGCCAACATCTAGCCGCTGTGACAGGCGCGCTACAG 1673

Db 601 TCCGGAGAGCTGTGAGGGGCCAACATCTAGCCGCTGTGACAGGCGCGCTACAG 660

QY 1674 ATGAATACCAAGCATCTGTGTACA-CCGATCTGATGTGTTGAGGCTGGGCGCTGCT 1732

Db 661 ATGAATACCAAGCATCTGTGTACA-CCGATCTGATGTGTTGAGGCTGGGCGCTGCT 720

QY 1733 CACATCTGCTCCTCCTCGGCAAGCCCGGCTTCCGCTACAGACTGTG 1785

Db 721 CACATCTGCTCCTCCTCGGCAAGCCCGGCTTCCGCTACAGACTGTG 773

RESULT 10  
B327597 1097 bp mRNA linear EST 27-FEB-2001  
LOCUS B327597 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:4564388 5',  
DEFINITION mRNA sequence.  
ACCESSION B327597  
VERSION B327597.1 GI:13134035  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
AUTHORS NIH-MGC Http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabe-remail.nih.gov  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: Incyte Genomics, Inc.  
DNA Sequencing: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:  
http://image.llnl.gov  
Plate: LICM1279 row: 1 column: 21  
High quality sequence stop: 783.  
Location/Qualifiers  
FEATURES  
source 1..1097  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4564388"  
/tissue\_type="renal cell adenocarcinoma"

```

/lab host="DH10B (phage-resistant)"
/clone lib="NH10B"
/notes="Organ: Kidney; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGAG(G). Size selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using Zap-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 212 a 340 c 338 g 205 t 2 others

Query Match 28.8%; Score 741.8; DB 10; Length 1097;  
Best Local Similarity 91.5%; Pred. No. 3.5e-129;  
Matches 843; Conservative 0; Mismatches 67; Indels 11; Gaps 5;

QY 967 CCATCATCGGAGGCTCTTCTCTCGAGAGCTGATCAAGATTGGACACACATCT 1026  
Db 2 CCATCATCGGAGGCTCTTCTCTCGAGAGCTGATCAAGATTGGACACACATCT 61

QY 1027 ATGGGGCCGACACTTTCATGATGATGAGGACCACTTCTCAGAGCCCTCTACCTTGGCG 1086  
Db 62 ATGGGGCCGACACTTTCATGATGATGAGGACCACTTCTCAGAGCCCTCTACCTTGGCG 121

QY 1087 CAGCCACACTGCGCTGATGAGGCGCATGACTGAGTGAAGTGAAGCTGTGTGCTGC 1146  
Db 122 CAGCCACACTGCGCTGATGAGGCGCATGACTGAGTGAAGTGAAGCTGTGTGCTGC 181

QY 1147 TCAGAGGCTGCTTCCACACACAGCGGAGTTCTGGGGGCCCGCCAGATCAGGGCTG 1206  
Db 182 TCAGAGGCTGCTTCCACACACAGCGGAGTTCTGGGGGCCCGCCAGATCAGGGCTG 241

QY 1207 TGGTGGAGCTGTGCCCCGCTGCGCCTCTCTGCTTGTGACCTGTTTGTGAGAGCAGC 1266  
Db 242 TGGTGGAGCTGTGCCCCGCTGCGCCTCTCTGCTTGTGAGCTGTTTGTGAGAGCAGC 301

QY 1267 CTGTGTATACCCGACAGCTCTCTCTCGAGGCGCAGCCCTTCACTGTGTGATGCTGACA 1326  
Db 302 CTGTGTATACCCGACAGCTCTCTCTCGAGGCGCAGCCCTTCACTGTGTGATGCTGACA 361

QY 1327 ACTTTGGGGGAAACATGCTTTTGTGAGCCTTGAAGGCTGTGAACGAGGCCGAGAAG 1386  
Db 362 ACTTTGGGGGAAACATGCTTTTGTGAGCCTTGAAGGCTGTGAACGAGGCCGAGAAG 421

QY 1387 CTGCGCGCTCTTCCCACTTCCACATGATGAGGACAGGSCATGSCCCCGAGGGCATCA 1446  
Db 422 CTGCGCGCTCTTCCCACTTCCACATGATGAGGACAGGSCATGSCCCCGAGGGCATCA 481

QY 1447 GCGGAAAGAGGTGCTATTTCCCTCACTGAGGCTGAGGCGGAGGAAAGAACCCAGTGC 1506  
Db 482 GCGGAAAGAGGTGCTATTTCCCTCACTGAGGCTGAGGCGGAGGAAAGAACCCAGTGC 541

QY 1507 CAGATTGGAGCCTGGGTGACCAAGCTTTGCCCGCGGATGAGGCTTCCACCCG 1566  
Db 542 CAGATTGGAGCCTGGGTGACCAAGCTTTGCCCGCGGATGAGGCTTCCACCCG 601

QY 1567 ACGAGAGGCGAGCTGAGAGCTACTGCTCGGAGTGTGTAACAATGCTCCGGGAGAGCT 1626  
Db 602 ACGAGAGGCGAGCTGAGAGCTACTGCTCGGAGTGTGTAACAATGCTCCGGGAGAGCT 661

QY 1627 GCAAGGGGCAAAATCGTAGCCGCTGCTGAGGCGCGCTTCCCTACAGATGATACAGCA 1686  
Db 662 GCAAGGGGCAAAATCGTAGCCGCTGCTGAGGCGCGCTTCCCTACAGATGATACAGCA 720

QY 1687 TCTGTGTAACCAATCTGATGTGTGTTGAGGCTGGCGGCTGCTCAATCTGCTCCCT 1746  
Db 721 TCTGTGTAACCAATCTGATGTGTGTTGAGGCTGGCGGCTGCTCAATCTGCTCCCT 779

QY 1747 CCTGGGCAACAGCCCGCTTCCGCTACAGACTGTGAGACTCACTCTGGCAGAGTGC 1806  
Db 780 GGC-----ACAACCGCGCTTGGGTA-GAAGGTGAGACTCACTCGGC--GAGGTC 830

QY 1807 AGAGCTGTGAGCTTGTACTATGAGAGGCAAGAGCCCTACTGAGCAAGAGCTGC 1866  
Db 831 AGAAGCGTACCTGTGAGCAATGAGAGAGGAGCAAAAGGCCAACCTGTGACAGGGCGGCG 890

QY 1867 CCTCCCTGTTGAGGCGCTTGAG 1887  
Db 891 CCCCCTTGAAGGCGCGGCGAG 911

## RESULT 11

BM976494/c

LOCUS

DEFINITION

UI-CF-EN1-acz-1-12-0-UI.61 UI-CF-EN1 Homo sapiens cDNA clone

BM976494

## ACCESSION

VERSION

BM976494.1 GI:19594088

EST

Homo sapiens (human)

758 bp mRNA linear EST 21-FEB-2003

## KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

AUTHORS

TITLE

Normalizaton and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

Contact: McCray, PB

## JOURNAL

MEDLINE

PUBMED

8889548

McCrack Lab

University of Iowa

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA

## COMMENT

Email: paul-mccray@iowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa

CDNA Library preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Genetics (www.resgen.com) or from Open Biosystems

Seg primer: M13 FORWARD

## FEATURES

Location/Qualifiers

1..758

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="UI-CF-EN1-acz-1-12-0-UI"

/tissue="Primary Lung Cystic Fibrosis Epithelial Cells"

## dev\_stage="Adult"

/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone\_lib="UI-CF-EN1"

/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;

UI-CF-EN1 is a normalized cDNA library containing the

following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to

## Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. First strand cDNA synthesis was primed with an

oligo-dT primer containing a Not I site. Double stranded

cDNA was ligated to an EcoR I adaptor, digested with Not

I, and cloned directionally into pT73-Pac vector. The

oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

## located between the Not I site and the Gp18 tail. The

sequence tag for this library is CTGCTCAGGT.

TAG Lib=UI-CF-EN1

TAG Tissue=Human Lung Epithelial Cell Lines untreated LPS

6hr to LPS 24h

TAG\_SEQ=CTGCTCAGGT"



[illegible]

RESULT	13
BG476070	
LOCUS	816 bp MENA linear EST 21-MAR-2001
DEFINITION	G08520954P1 NIH_MGC_20 Homo sapiens CDNA clone IMAGE:463620 5', mRNA sequence.
ACCESSION	BG476070
VERSION	BG476070.1 GI:13408349
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE  
1 (bases 1 to 816)  
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC).  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: [CGABpds-remail.nhl.nih.gov](mailto:CGABpds-remail.nhl.nih.gov)  
 Tissue Procurement: [ATCCC/DCD/Drp](http://ATCCC/DCD/Drp)  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrived by: The I.M.A.G.E. Consortium (LINTL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINTL at:  
<http://image.llnl.gov>  
 Plate: LHCWL403 row: k column: 13  
 High quality sequence row: 778.

FEATURES  
source

Location/Qualifiers  
1. .816

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4639620"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_id="NH MGC 20"
/note="Organ: skin; Vector: pOT8; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT
ORIGIN
162 a
241 c 256 g 157 t

```

Query Match	27.5%	Score	707.6	DB	10	Length	816
Best Local Similarity	97.3%	Pred. No.	9	5=123			
Matches	794	Conservative	0	Mismatches	14	Indels	8
Gaps							7

  

QY	1682	CAGCATCTGTTCACACCGATCTGATGTGTTTAAAGGCTTGCGGCGCTGCTCAACATTCGC	1744
DB	2	CAGCATCTGTTCACACCGATCTGATGTGTTTAAAGGCTTGCGGCGCTGCTCAACATTCGC	61

[illegible]

RESULT 14	
BQ772664/c	
LOCUS	BQ772664
DEFINITION	707 bp mRNA linear EST 26-JUL-2002 ui-H-FEO-bhg-p-12-O-ui-s1 NC1 CGAP FEO Homo sapiens cDNA clone ut-H-FEO-bhg-p-12-O-ut-37 RNA-seq

ACCESSION	U174750.1
VERSION	BQ772664
KEYWORDS	BQ772664.1
SOURCE	EST
ORGANISM	Homo sapiens (human)

REFERENCE  
1 (bases 1 to 707)  
Euariyocota; Metazoa; Chirocata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

**AUTHORS** NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
**JOURNAL** Unpublished  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: James Martin  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@uiowa.edu  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

# FEATURES

source  
 1..707  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-FEO-dbg-p-12-0-UI"  
 /tissue\_type="Chondrosarcoma Cell line"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_id="NCI CGAP\_FEO"  
 /note="Vector: pT73-Pac (Pharmacia) with a modified  
 polylinker. Site 1: EcoR I; Site 2: Not I; NCI CGAP\_FEO is  
 a cDNA library containing the following tissue(s): a pool  
 of 3 chondrosarcoma cell lines (grade 2) The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor, digested with Not I, and cloned directionally  
 into pT73-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is  
 CGCTACGAC. The cell lines was provided by Dr James Martin  
 of University of Iowa.  
 TAG LIB=UI-H-FEO  
 TAG TISSUE=Human grade 2 chondrosarcoma cell line pool  
 TAG\_SEQ=CGTACGAC"  
 BASE COUNT 135 a 214 c 198 g 158 t 2 others  
 ORIGIN  
 Query Match 27.3%; Score 702; DB 13; Length 707;  
 Best Local Similarity 99.7%; Pred.No.1.1e-121;  
 Matches 702; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 1853 GAGCAGAGAGAGTGGCTCCCTGTTAGAGGCTGAGAGCGCTCGGCTATGAGCTCTGCC 1912  
 707 GAGCAGAGAGAGTGGCTCCCTGTTAGAGGCTGAGAGCGCTCGGCTATGAGCTCTGCC 648  
 1913 GGCACCTGACGAGAGTCTGCTAGTACAGCGCTTTCTTCTGCTGCTGCTAGACCA 1972  
 647 GGCACCTGACGAGAGTCTGCTAGTACAGCGCTTTCTTCTGCTGCTGCTAGACCA 588  
 1973 GGGCCGAGAGAGGCGGACGAGTGGGCGAGGCGCATTTCTAGAGCAGAAACAGCGGCTA 2032  
 587 GGGCCGAGAGAGGCGGACGAGTGGGCGAGGCGCATTTCTAGAGCAGAAACAGCGGCTA 528  
 2033 CCAGCTGACCTTGTGGGCGGCGAAGAGCAACTCTGACCTATGCGCAACAGCACTGGC 2092  
 527 CCAGCTGACCTTGTGGGCGGCGAAGAGCAACTCTGACCTATGCGCAACAGCACTGGC 468  
 2093 GGGGTTGGTGGCACTACTACACCCCTGCTGGGCGCTTTTCTGAGAGCGCTGTTGA 2152  
 467 GGGGTTGGTGGCACTACTACACCCCTGCTGGGCGCTTTTCTGAGAGCGCTGTTGA 408  
 2153 CAGTGTGGCCAGAGGAGATCCCTTCCAGAGCAGCACTTGAACAAATGTCTTCAACT 2212  
 407 CAGTGTGGCCAGAGGAGATCCCTTCCAGAGCAGCACTTGAACAAATGTCTTCAACT 348  
 2213 GAGAGCGAGCTTGTCTTCTAGCAGAGAGTACCCACAGCGCGGAGAGAGACACT 2272  
 347 GAGAGCGAGCTTGTCTTCTAGCAGAGAGTACCCACAGCGCGGAGAGAGACACT 288  
 2273 GAGAGCGAGCTTGTCTTCTAGCAGAGAGTACCCACAGCGCGGAGAGAGACT 2332

Db 287 GAGCTGGCCAGAGAGATCTTCTCAATATTATCCCGGCTGGGCGGCTCTTGCTG 228  
 QY 2333 ATGATTCGCCACACACTGAGGCTGTTTCCGTAATTCACAGGAGATTCCAGGCGCCA 2392  
 Db 227 ATGATTCGCCACACACTGAGGCTGTTTCCGTAATTCACAGGAGATTCCAGGCGCCA 168  
 QY 2393 GAGCTGGACAGATCAGATCAGATTAACCCAGGCTTGGAGAGAGCCCGCAGGCTGTGT 2452  
 Db 167 GAGCTGGACAGATCAGATCAGATTAACCCAGGCTTGGAGAGAGCCCGCAGGCTGTGT 108  
 QY 2453 GGGGCTGACCTGGGGGGATTGGAGGAAATGACCGCTCCACACACCAACCAAGTGT 2512  
 Db 107 GGGGCTGACCTGGGGGGATTGGAGGAAATGACCGCTCCACACACCAACCAAGTGT 48  
 QY 2513 GGGATTAAAGTACTGTTTCTTCCACTTAAACCAACCAACCAACCAACCAACCAACCA 2556  
 Db 47 GGGATTAAAGTACTGTTTCTTCCACTTAAACCAACCAACCAACCAACCAACCAACCA 4  
 RESULT 15  
 BQ006037/c 707 bp mRNA linear EST 26-MAR-2002  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 707)  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Dr. Jose Mercuende  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be found  
 through the I.M.A.G.E. Consortium/HLN at: <http://image.llnl.gov>  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

## FEATURES

source  
 1..707  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5845510"  
 /tissue\_type="Chondrosarcoma"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_id="NCI CGAP\_E11"  
 /note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)  
 with a modified polylinker. Site 1: EcoR I; Site 2: Not I;  
 NCI CGAP E11 is a normalized cDNA library containing the  
 following tissue(s): Chondrosarcoma. The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor, digested with Not I, and cloned directionally  
 into pT73-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is  
 ACACCTTGAC.  
 TAG LIB=UI-H-E11  
 TAG TISSUE=chondrosarcoma  
 TAG\_SEQ=ACACCTTGAC"

BASE COUNT 136 a 213 c 199 g 158 t 1 others  
ORIGIN

Query Match 27.2%; Score 701.4; DB 12; Length 707;  
Best Local Similarity 99.7%; Pred. No. 1.4e-121;  
Matches 702; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1853 GAGCAGAGAGCTGGCTCCCTGTTGAGGGCTGAGAGCGCTGCTTGAAGCTGCTGCC 1912
Db 707 GAGCAAGAGAGCTGCCCTCCCTGTTGAGGGCTGAGAGCGCTGCTTGAAGCTGCTGCC 648
QY 1913 GGCACCTGACGAGAGCTGGCTGGTAGACAGCCGCTTTGCTGGGAGCTGGTAGACA 1972
Db 647 GGCACCTGACGAGAGCTGGCTGGTAGACAGCCGCTTTGCTGGGAGCTGGTAGACA 588
QY 1973 GGGCCGAGAGAGGGGAGTAGAGGCGAGGCGATTCTACGAGCAGAAACAGCGGCTA 2032
Db 587 GGGCCGAGAGAGGGGAGTAGAGGCGAGGCGATTCTACGAGCAGAAACAGCGGCTA 528
QY 2033 CCAGCTGACCTTGTGGGGGCGAAGAGGCAACATCCTGACATATGCCAACAGCAGCTGGC 2092
Db 527 CCAGCTGACCTTGTGGGGGCGAAGAGGCAACATCCTGACATATGCCAACAGCAGCTGGC 468
QY 2093 GGGGTTGGTGGGCACTACTACACCCCTGGCTGGGGGCTTTTCCCTGAGAGCGCTGGTTGA 2152
Db 467 GGGGTTGGTGGGCACTACTACACCCCTGGCTGGGGGCTTTTCCCTGAGAGCGCTGGTTGA 408
QY 2153 CAGTGTGGGCGCAGAGGAGATCCCTTCCACAGCAGCAGATTTGACAAAAATGTTCCAACT 2212
Db 407 CAGTGTGGGCGCAGAGGAGATCCCTTCCACAGCAGCAGATTTGACAAAAATGTTCCAACT 348
QY 2213 GAGAGAGGCTTGTGTTCTTACAGCAGAGGTAACCCAGCCAGCCGCGAGAGACACTGT 2272
Db 347 GAGAGAGGCTTGTGTTCTTACAGCAGAGGTAACCCAGCCAGCCGCGAGAGACACTGT 288
QY 2273 GGAACCTGGCCAGAGAAATCTTCTCAAAATATTACCCGGCTGGGTGGCCGGCTCTTGGTG 2332
Db 287 GGAACCTGGCCAGAGAAATCTTCTCAAAATATTACCCGGCTGGGTGGCCGGCTCTTGGTG 228
QY 2333 ATAGATTGCGCACCACTGGGGCTTGTTCGCTAATTCAGGGGCAATTCAGGGGCCCA 2392
Db 227 ATAGATTGCGCACCACTGGGGCTTGTTCGCTAATTCAGGGGCAATTCAGGGGCCCA 168
QY 2393 GAGCTGGACAGACATCAGAGATTAACCCAGGCTGGGAGAGAGGCCCAAGGCTGTGAT 2452
Db 167 GAGCTGGACAGACATCAGAGATTAACCCAGGCTGGGAGAGAGGCCCAAGGCTGTGAT 108
QY 2453 GGGGCTGACCTGGGGGGAATTGAGGGAAATGACCTGCCCTCCACACCAACCCAAAGTGT 2512
Db 107 GGGGCTGACCTGGGGGGAATTGAGGGAAATGACCTGCCCTCCACACCAACCCAAAGTGT 48
QY 2513 GGGATTAAAGTACTGTTTCTTCCACTTAAAAAATAAAAA 2556
Db 47 GGGATTAAAGTACTGTTTCTTCCACTTAAAAAATAAAAA 4
```

Search completed: February 13, 2004, 13:31:55  
Job time : 5311 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: February 13, 2004, 09:08:09 ; Search time 150 seconds  
(without alignments)  
7577.077 Million cell updates/sec

Title: US-09-836-613-1  
Perfect score: 2575  
Sequence: 1 CCGGGCTTACCTTCGGGT.....AGTCAGCGCGCGCAATTC 2575

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	2575	100.0	2575 3	US-09-077-354B-1
2	1426.6	55.4	10380 3	US-09-077-354B-3
3	70.4	2.7	5173 1	US-08-242-677-1
4	66.2	2.6	1288 1	US-08-440-856A-9
5	66.2	2.6	152331 3	US-09-128-155-16
6	62.2	2.4	4411529 3	US-08-103-840A-1
7	61	2.4	4403765 3	US-09-103-840A-2
8	60.2	2.3	4403765 3	US-09-103-840A-2
9	60.2	2.3	4411529 3	US-09-103-840A-1
10	59.6	2.3	4257 2	US-08-690-473-1
11	59.6	2.3	4257 3	US-09-259-821A-1
12	59.6	2.3	4257 3	US-08-843-659-1
13	59.6	2.3	12001 1	US-08-458-568A-11
14	58.6	2.3	1028 3	US-08-118-200-1
15	58.6	2.3	1028 3	US-08-458-745-1
16	58.2	2.3	1236 1	US-08-440-856A-1
17	58.2	2.3	471 4	US-09-252-991A-12181
18	58.2	2.3	531 4	US-09-252-991A-12413
19	58.2	2.3	567 4	US-09-252-991A-12364
20	58.2	2.3	747 4	US-09-252-991A-12243
21	58.2	2.3	2052 4	US-09-252-991A-12103
22	58.2	2.3	15378 3	US-08-785-420-1
23	57.6	2.2	4776 2	US-08-852-401-1
24	57.4	2.2	525 4	US-09-252-991A-14501
25	57.4	2.2	1092 4	US-09-252-991A-14759
26	57.4	2.2	2145 4	US-09-252-991A-14638
27	57.4	2.2	2487 4	US-09-252-991A-14985

28	57	2.2	44377 2	US-08-804-227C-7	Sequence 7, Appl
29	57	2.2	44377 2	US-08-804-198-1	Sequence 1, Appl
30	56	2.2	44377 2	US-08-804-227C-7	Sequence 7, Appl
31	56	2.2	44377 2	US-08-804-198-1	Sequence 1, Appl
32	55.8	2.2	47981 4	US-09-679-279-1	Sequence 1, Appl
33	55.2	2.1	34094 4	US-09-252-034-1	Sequence 1, Appl
34	55	2.1	2481 4	US-09-894-998A-35	Sequence 35, Appl
35	55	2.1	2721 6	5215881-2	Patent No. 5215881
36	55	2.1	4257 2	US-08-690-473-1	Sequence 1, Appl
37	55	2.1	4257 3	US-09-259-821A-1	Sequence 1, Appl
38	55	2.1	4257 3	US-08-843-659-1	Sequence 1, Appl
39	55	2.1	12001 1	US-08-458-568A-11	Sequence 11, Appl
40	54.8	2.1	3946 3	US-09-083-351-1	Sequence 1, Appl
41	54.8	2.1	3946 3	US-09-083-352-1	Sequence 1, Appl
42	54.6	2.1	396 4	US-09-416-509C-2	Sequence 2, Appl
43	54.6	2.1	2189 4	US-08-846-020A-1	Sequence 1, Appl
44	54.6	2.1	2189 4	US-09-617-871-1	Sequence 1, Appl
45	54.4	2.1	2312 1	US-07-736-178C-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-077-354B-1  
Sequence 1, Application US/09077354B  
Patent No. 6255096  
GENERAL INFORMATION:  
APPLICANT: HOBWOOD, JOHN JOSEPH; SCOTT, HANISH STEELE;  
APPLICANT: WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART  
TITLE OF INVENTION: SYNTHETIC MAMMALIAN  
TITLE OF INVENTION: N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSER: SCULLY, SCOTT, MURPHY & PRESSER  
STREET: 400 GARDEN CITY PLAZA  
CITY: GARDEN CITY  
STATE: NEW YORK  
COUNTRY: UNITED STATES  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/077,354B  
FILING DATE: 22-APRIL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00747  
FILING DATE: 22-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: POKALSKY, ANN R.  
REGISTRATION NUMBER: 34,697  
REFERENCE/DOCKET NUMBER: 12416  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516 742 4343  
TELEFAX: 516 742 4366  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2575 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Peripheral Blood  
CELL TYPE: Leukocyte  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 102..2330  
US-09-077-354B-1

Query Match 100.0%; Score 2575; DB 3; Length 2575;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CCCGGGCTTAGCTTCCGAGTCCAGTGGCCGAGCCCGCAGCTGATTGAAACGGGGCCG 60  
 Db 1 CCCGGGCTTAGCTTCCGAGTCCAGTGGCCGAGCCCGCAGCTGATTGAAACGGGGCCG 60

61 CCCCAACCCCTGGCCGTGGCGGAGACCCCGAGGACTGAGACCATGAGAGCGGGCGGTG 120  
 Db 61 CCCCAACCCCTGGCCGTGGCGGAGACCCCGAGGACTGAGACCATGAGAGCGGGCGGTG 120

61 CCCCAACCCCTGGCCGTGGCGGAGACCCCGAGGACTGAGACCATGAGAGCGGGCGGTG 120  
 Db 61 CCCCAACCCCTGGCCGTGGCGGAGACCCCGAGGACTGAGACCATGAGAGCGGGCGGTG 120

121 CCGCGCGGCTGGGGGCTCTTCTCTGCGCGGGCCCGGGCGCGAGCGCAAGAGCC 180  
 Db 121 CCGCGCGGCTGGGGGCTCTTCTCTGCGCGGGCCCGGGCGCGAGCGCAAGAGCC 180

181 GGGAGGCGGCGCGCGCTGCGGGCGCTGCGCGCGCTGCGGGCGAGGCGCGCGCG 240  
 Db 181 GGGAGGCGGCGCGCGCTGCGGGCGCTGCGCGCGCTGCGGGCGAGGCGCGCGCG 240

241 ACTTCTCCGCTGCTGAGAGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCG 300  
 Db 241 ACTTCTCCGCTGCTGAGAGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCG 300

301 TGGGCGGCGGCG 360  
 Db 301 TGGGCGGCGGCG 360

361 CCGGGCTGAGACCGCTACCTGCGGAGCTTCTGAGCTGAGCAAGTGGCTGCGCGCTG 420  
 Db 361 CCGGGCTGAGACCGCTACCTGCGGAGCTTCTGAGCTGAGCAAGTGGCTGCGCGCTG 420

421 AGCTGGGCTGCG 480  
 Db 421 AGCTGGGCTGCG 480

481 ACGAGTACCGCTATTTCAGAGATGCTGAGACGAAAGCTACTCTTCGCTGCTGAGAG 540  
 Db 481 ACGAGTACCGCTATTTCAGAGATGCTGAGACGAAAGCTACTCTTCGCTGCTGAGAG 540

541 GGGCGCGCTGGAGCGAGAGATGAGCTGAGTGGCGCTGAGTGGCATCAACTGGCACTG 600  
 Db 541 GGGCGCGCTGGAGCGAGAGATGAGCTGAGTGGCGCTGAGTGGCATCAACTGGCACTG 600

601 CCGAGGCGGCGCGAGAGGCGCATCTGGAGCGGGTGAACCTGGCTTGGGCGTGAACCG 660  
 Db 601 CCGAGGCGGCGCGAGAGGCGCATCTGGAGCGGGTGAACCTGGCTTGGGCGTGAACCG 660

661 CAGAGTCAATGAGTCTTACTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 720  
 Db 661 CAGAGTCAATGAGTCTTACTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 720

721 TGGACACTTGGAGTGGCCCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 780  
 Db 721 TGGACACTTGGAGTGGCCCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 780

781 ACCGGGCTGGAGCGAGATGCGCTCTTGGCATGACCCAGTGGCTGAGATTGGCGG 840  
 Db 781 ACCGGGCTGGAGCGAGATGCGCTCTTGGCATGACCCAGTGGCTGAGATTGGCGG 840

841 GGCATGTTCCGAGGCTGTACCAAGGCTGTTCCCTCAGTCAATGTCAGAGAGTGGGA 900  
 Db 841 GGCATGTTCCGAGGCTGTACCAAGGCTGTTCCCTCAGTCAATGTCAGAGAGTGGGA 900

901 GTTGGGGCACTTAACTGTTCTACTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 960  
 Db 901 GTTGGGGCACTTAACTGTTCTACTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 960

961 TATTCCTCATCTGGAGGCTCTTCTTCTGAGAGCTGATCAAGAGTGGGACAGACC 1020  
 Db 961 TATTCCTCATCTGGAGGCTCTTCTTCTGAGAGCTGATCAAGAGTGGGACAGACC 1020

1021 ACATCTATGGGGCGGACACTTTCATGAGATGAGCCACCTTCTCTAGAGCCCTCTTACC 1080  
 Db 1021 ACATCTATGGGGCGGACACTTTCATGAGATGAGCCACCTTCTCTAGAGCCCTCTTACC 1080

1081 TTGGCCGAGCGACCATGCGCGTATGAGGCGCATGCTGACAGGATACTGAGAGCTGTGT 1140  
 Db 1081 TTGGCCGAGCGACCATGCGCGTATGAGGCGCATGCTGACAGGATACTGAGAGCTGTGT 1140

1141 GGCCTGCTCCAGAGCTGGCTTCTTCAGACACAGCCGAGTTCGGGGGCGCCCGCAGATCA 1200  
 Db 1141 GGCCTGCTCCAGAGCTGGCTTCTTCAGACACAGCCGAGTTCGGGGGCGCCCGCAGATCA 1200

1201 GGGCTGTGCTGGAGAGCTGTGCCCCCGTGGCCGCTTCTGAGCTGTTGCTGAGA 1260  
 Db 1201 GGGCTGTGCTGGAGAGCTGTGCCCCCGTGGCCGCTTCTGAGCTGTTGCTGAGA 1260

1261 GCGAGCTGTGTATACCCCGCATGCTCTTCCAGGCGCAGCCCTTCATCTGGTGATGC 1320  
 Db 1261 GCGAGCTGTGTATACCCCGCATGCTCTTCCAGGCGCAGCCCTTCATCTGGTGATGC 1320

1321 TGGCAAACTTTGGGGGAAACATGCTCTTTTGGAGCCCTTGAAGGCTGGAACGAGGCG 1380  
 Db 1321 TGGCAAACTTTGGGGGAAACATGCTCTTTTGGAGCCCTTGAAGGCTGGAACGAGGCG 1380

1381 CAGAGGCTGCCGCTCTTCCCAACTCCACATGCTAGGCAAGGCGATGCGCCCGAGAG 1440  
 Db 1381 CAGAGGCTGCCGCTCTTCCCAACTCCACATGCTAGGCAAGGCGATGCGCCCGAGAG 1440

1441 GCATGAGCCAGAAAGAAAGTGTCTATTCCCTCATGCTGAGCTGGGCTGGCAAGAGAG 1500  
 Db 1441 GCATGAGCCAGAAAGAAAGTGTCTATTCCCTCATGCTGAGCTGGGCTGGCAAGAGAG 1500

1501 CAGTGCAGATTGGGAGGCTGGGAGACAGCTTGGCGCGCGCGGATGAGGAGTCTCCC 1560  
 Db 1501 CAGTGCAGATTGGGAGGCTGGGAGACAGCTTGGCGCGCGCGGATGAGGAGTCTCCC 1560

1561 ACCCGAGCGAGGGGAGCGCTGAGAGGCTACTGCTCGGAGTGTGTAACAATGCTCGGGG 1620  
 Db 1561 ACCCGAGCGAGGGGAGCGCTGAGAGGCTACTGCTCGGAGTGTGTAACAATGCTCGGGG 1620

1621 AGGCTGCAAGGGGCAAAATGTAGAGCCGCTGAGGCGGCGCTCCCTACAGATGAATA 1680  
 Db 1621 AGGCTGCAAGGGGCAAAATGTAGAGCCGCTGAGGCGGCGCTCCCTACAGATGAATA 1680

1681 CCAAGCATCTGGTACCAACCAATCTGATGTGTTGAGAGCTGCGGCTGCTCAATCTG 1740  
 Db 1681 CCAAGCATCTGGTACCAACCAATCTGATGTGTTGAGAGCTGCGGCTGCTCAATCTG 1740

1741 CTCCTCTCCCTGGCCACAGCCCGGCTTCCGCTACAGACCTGAGACTCGGAGG 1800  
 Db 1741 CTCCTCTCCCTGGCCACAGCCCGGCTTCCGCTACAGACCTGAGACTCGGAGG 1800

1801 CAGTGCAGAGAGCTGTGAGCTTGTACTATGAGAGGCGCAAGAGCGCTTACTGAGCAAG 1860  
 Db 1801 CAGTGCAGAGAGCTGTGAGCTTGTACTATGAGAGGCGCAAGAGCGCTTACTGAGCAAG 1860

1861 AGCTGGCTCTCTTGTGAGGCTGAGAGGCTGCTGAGCTTATGAGTCTTCCCGCATGG 1920  
 Db 1861 AGCTGGCTCTCTTGTGAGGCTGAGAGGCTGCTGAGCTTATGAGTCTTCCCGCATGG 1920

1921 AGCTGGCTCTCTTGTGAGGCTGAGAGGCTGCTGAGCTTATGAGTCTTCCCGCATGG 1980  
 Db 1921 AGCTGGCTCTCTTGTGAGGCTGAGAGGCTGCTGAGCTTATGAGTCTTCCCGCATGG 1980

1981 CAGCGGCACTGATGAGGCGGCGGATTTCTACAGAGCAAGAGCGGCTTACCAAGTGA 2040  
 Db 1981 CAGCGGCACTGATGAGGCGGCGGATTTCTACAGAGCAAGAGCGGCTTACCAAGTGA 2040

2041 CTTGTGGGGGCGAGAGGCAACATCTGAGCTATGCAACAAGCACTGGCGGGTGG 2100  
 Db 2041 CTTGTGGGGGCGAGAGGCAACATCTGAGCTATGCAACAAGCACTGGCGGGTGG 2100

2101 TGGCCAACTACTACACCCCTGCGCGGCTTTTCTGAGAGGCGCTGTGACAGTGTGG 2160

Db 2101 TGGGCACTACTACACCCCTGCTGGCGGCTTTCTTGAGAGCGCTGTGACAGTGTGG 2160  
Qy 2161 CCCAGGGGATCCCTTTCCACAGACAGATTGCAAAATGTCTTCCACTGAGAGG 2220  
Db 2161 CCCAGGGGATCCCTTTCCACAGACAGATTGCAAAATGTCTTCCACTGAGAGG 2220  
Qy 2221 CTTTCGTTCTCAGCAGACAGAGTACCAGCCAGCCGAGAGAGACACTGTGAGCTGG 2280  
Db 2221 CTTTCGTTCTCAGCAGAGAGTACCAGCCAGCCGAGAGAGACACTGTGAGCTGG 2280  
Qy 2281 CCAAGAGATCTTCTCTAATATTAACCCGCTGGGTGGGCGGCTCTGTGTATGATTTC 2340  
Db 2281 CCAAGAGATCTTCTCTAATATTAACCCGCTGGGTGGGCGGCTCTGTGTATGATTTC 2340  
Qy 2341 GCCACACTGGGCTTGTCTTCCGCTAATTCAGGCGAGATTCCAGGCGCCAGAGCTGGA 2400  
Db 2341 GCCACACTGGGCTTGTCTTCCGCTAATTCAGGCGAGATTCCAGGCGCCAGAGCTGGA 2400  
Qy 2401 CAGACATCACAGATTAACCCAGGCTGGAGAGAGCCCAAGGCTGTGTGTGTTG 2460  
Db 2401 CAGACATCACAGATTAACCCAGGCTGGAGAGAGCCCAAGGCTGTGTGTGTTG 2460  
Qy 2461 ACCGCGGGGAGTTGAGAGGAAATGACCTGCTTCCACACACCCAAAGTGGGATTAA 2520  
Db 2461 ACCGCGGGGAGTTGAGAGGAAATGACCTGCTTCCACACACCCAAAGTGGGATTAA 2520  
Qy 2521 AGTACTGTTTCTTCCACTTAAAAAAGTCGAGCGCCGCAATTTC 2575  
Db 2521 AGTACTGTTTCTTCCACTTAAAAAAGTCGAGCGCCGCAATTTC 2575

## RESULT 2

US-09-077-354B-3  
Sequence 3, Application US/09077354B  
Patent No. 6255096  
GENERAL INFORMATION:  
APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;  
APPLICANT: WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART  
TITLE OF INVENTION: SYNTHETIC NANNALINAN  
TITLE OF INVENTION: '-N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER  
STREET: 400 GARDEN CITY PLAZA  
CITY: GARDEN CITY  
STATE: NEW YORK  
COUNTRY: UNITED STATES  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/077,354B  
FILING DATE: 22-APRIL-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00747  
FILING DATE: 22-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: POKALSKY, ANN R.  
REGISTRATION NUMBER: 34,697  
REFERENCE/DOCKET NUMBER: 12416  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516 742 4343  
TELEFAX: 516 742 4366  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10380 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: Chromosome 17  
FEATURE:  
NAME/KEY: exon 1  
LOCATION: 990..1372  
FEATURE:  
NAME/KEY: exon 2  
LOCATION: 2115..2262  
FEATURE:  
NAME/KEY: exon 3  
LOCATION: 3056..3202  
FEATURE:  
NAME/KEY: exon 4  
LOCATION: 3387..3472  
FEATURE:  
NAME/KEY: exon 5  
LOCATION: 5667..5823  
FEATURE:  
NAME/KEY: exon 6  
LOCATION: 7745..8955  
US-09-077-354B-3  
Query Match 55.4%; Score 1426.6; DB 3; Length 10380;  
Best Local Similarity 99.7%; Pred. No. 3.3e-289;  
Matches 1429; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1112 CATGACTCAGATGATGATGAGGCTGTGTGCTGCTCCAGAGCTGCTCTTCCAGACCA 1171  
Db 7734 CCGTTCACAGAGGATATGAGGCTGTGTGCTGCTCCAGAGCTGCTCTTCCAGACCA 7793  
Qy 1172 GCCGAGATTCTGGGGGCCCGCCAGATCAGGGCTGTGCTGGAGCTGTGCCCTGGCG 1231  
Db 7794 GCCGAGATTCTGGGGGCCCGCCAGATCAGGGCTGTGCTGGAGCTGTGCCCTGGCG 7853  
Qy 1232 CTTCTGTTCTGAGACTGTTTGTGAGAGCAGGCTGTGATACCGGCACTGCTCCTT 1291  
Db 7854 CTTCTGTTCTGAGACTGTTTGTGAGAGCAGGCTGTGATACCGGCACTGCTCCTT 7913  
Qy 1292 CCAGGGCCAGCCCTTCATCTGTGTGATGCTGATGCAACTTTGGGGGAAACATGATCTTT 1351  
Db 7914 CCAGGGCCAGCCCTTCATCTGTGTGATGCTGATGCAACTTTGGGGGAAACATGATCTTT 7973  
Qy 1352 TGGAGCCCTAGAGGCTGTGAAGAGAGGCGCCAGAGCTGCCGCTTCCCACTCCAC 1411  
Db 7974 TGGAGCCCTAGAGGCTGTGAAGAGAGGCGCCAGAGCTGCCGCTTCCCACTCCAC 8033  
Qy 1412 CATGTAGGCAAGGCAATGAGCCCGGAGGAGCATGACCCAGAACGAGTGTCTATTCCT 1471  
Db 8034 CATGTAGGCAAGGCAATGAGCCCGGAGGAGCATGACCCAGAACGAGTGTCTATTCCT 8093  
Qy 1472 CATGCTAGCTGGGCTGGGCAAGAGACCCAGTGCATATTTGGAGCTGGGTACAG 1531  
Db 8094 CATGCTAGCTGGGCTGGGCAAGAGACCCAGTGCATATTTGGAGCTGGGTACAG 8153  
Qy 1532 CTTGCCCGCCGCGGATGAGGCTTCCACCCGAGCAGAGGGGAGCGTGGAGGCTACT 1591  
Db 8154 CTTGCCCGCCGCGGATGAGGCTTCCACCCGAGCAGAGGGGAGCGTGGAGGCTACT 8213  
Qy 1592 GCTCGGAGTGTGTACAACTGCTCCGGGAGAGCTGACAGGGGCCAATCTGAGCCGCT 1651  
Db 8214 GCTCGGAGTGTGTACAACTGCTCCGGGAGAGCTGACAGGGGCCAATCTGAGCCGCT 8273  
Qy 1652 GGTCAAGGGCGCGTCCCTACAGATGATATACAGACATCTGTATCAACCATCTGATGT 1711  
Db 8274 GGTCAAGGGCGCGTCCCTACAGATGATATACAGACATCTGTATCAACCATCTGATGT 8333  
Qy 1712 TAGAGCTTGGCGGCTGTGCTCACTGTGCTCCCTCCCTGGGCAACGAGCCGCTTCCG 1771  
Db 8334 TAGAGCTTGGCGGCTGTGCTCACTGTGCTCCCTCCCTGGGCAACGAGCCGCTTCCG 8393

OY	1772	CTACAGACTCTGGA	CTCACTGGGACAGTGCAGAGCTGCTGACCTGTACTAGA	1831	
Db	8394	CTACAGACTGCTGGA	CTCACTGGGACAGTGCAGAGCTGCTGACCTGTACTAGA	8453	
OY	1832	GGAGGCAGAGAGCGCCTA	CTGTAGCAAGAGCTGCGCTCCCTGTTGAGGGCTGAGGGCGT	1891	
Db	8454	GGAGGCAGAGAGCGCCTA	CTGTAGCAAGAGCTGCGCTCCCTGTTGAGGGCTGAGGGCGT	8513	
OY	1892	CCCTGGCCTATAGAGTG	CTGCGGCACTGGACAGAGTGTGGCTGTAGTACAGCGCGTTCTT	1951	
Db	8514	CCCTGGCCTATAGAGTG	CTGCGGCACTGGACAGAGTGTGGCTGTAGTACAGCGCGTTCTT	8573	
OY	1952	GCTGGGCACTGTGCTAG	AGCAGGCCGACAGCTGGGCACTCACTGATGAGCCGAGGCCGATT	2011	
Db	8574	GCTGGGCACTGTGCTAG	AGCAGGCCGACAGCTGGGCACTCACTGATGAGCCGAGGCCGATT	8633	
OY	2012	CTACGACACAGACAGCGCGT	ACAGCTGACCTGTTGGGGGACAGAAAGGCAACATCCTGGA	2071	
Db	8634	CTACGACACAGACAGCGCGT	ACAGCTGACCTGTTGGGGGACAGAAAGGCAACATCCTGGA	8693	
OY	2072	CTATGCCAACAGCAGCT	GGCGGTTGTTGGTGCACACTACACCCCTCGCTGGCGCT	2131	
Db	8694	CTATGCCAACAGCAGCT	GGCGGTTGTTGGTGCACACTACACCCCTCGCTGGCGCT	8753	
OY	2132	TTTCTCTGAGCGCTGTG	TGACAGTGGGCCAGGCACTCCCTTCCACACAGCACAGTT	2191	
Db	8754	TTTCTCTGAGCGCTGTG	TGACAGTGGGCCAGGCACTCCCTTCCACACAGCACAGTT	8813	
OY	2192	TGACAAAATGTCTTCCA	CTGAGCAGGCGTTGTTCTGACAGACAGAGGTACCCGAG	2251	
Db	8814	TGACAAAATGTCTTCCA	CTGAGCAGGCGTTGTTCTGACAGACAGAGGTACCCGAG	8873	
OY	2252	CCAGCCGCGAGAGACACT	GTGACCTGGGCCAGAGATCTTCCTCAATATTACCCGAG	2311	
Db	8874	CCAGCCGCGAGAGACACT	GTGACCTGGGCCAGAGATCTTCCTCAATATTACCCGAG	8933	
OY	2312	CTGGGTGGCGCGCTCTT	TGTGATGATTCGCACCACTGGGCGCTTGTTCCTCCGCTAATTC	2371	
Db	8934	CTGGGTGGCGCGCTCTT	TGTGATGATTCGCACCACTGGGCGCTTGTTCCTCCGCTAATTC	8993	
OY	2372	CAGGGCAGATTCCAGGGG	CCAGAGCTGGA	CAGACATCACAGGATAACCCAGGCTGGGAG	2431
Db	8994	CAGGGCAGATTCCAGGGG	CCAGAGCTGGA	CAGACATCACAGGATAACCCAGGCTGGGAG	9053
OY	2432	GAGGCCCACTGGCTGT	GTGGGTGACCTTGGGGGGAATTGGAGGGAAATGACTGCTGCC	2491	
Db	9054	GAGGCCCACTGGCTGT	GTGGGTGACCTTGGGGGGAATTGGAGGGAAATGACTGCTGCC	9113	
OY	2492	CTCCACACACACCCAA	AGTGTGGGATTTAAAGTACTGTTTCTTCCACTTAA	2544	
Db	9114	CTCCACACACACCCAA	AGTGTGGGATTTAAAGTACTGTTTCTTCCACTTAA	9166	

```

1      RESULT 3
2      US-08-24d-677-1
3      ; Sequence 1, Application US/08242677
4      ; Patent No. 5677143
5      ;
6      ; GENERAL INFORMATION:
7      ;
8      ; APPLICANT: Gaynor, Richard B
9      ; APPLICANT: Wu, Foon W.
10     ; TITLE OF INVENTION: Cellular Nucleic Acid Binding Protein
11     ; TITLE OF INVENTION: and Uses Thereof in regulating Gene Expression and in the
12     ; TITLE OF INVENTION: Treatment of AIDS
13     ; NUMBER OF SEQUENCES: 9
14     ; CORRESPONDENCE ADDRESS:
15     ; ADDRESSEE: Arnold, White & Durkee
16     ; STREET: P.O. Box 4433
17     ; CITY: Houston
18     ; STATE: TX
19     ; COUNTRY: USA
20     ; ZIP: 77210
21     ;
22     ; COMPUTER READABLE FORM:
23     ; MEDIUM TYPE: Floppy disk
24     ;

```

```

? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/242,677
? FILING DATE:
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Mayfield, Denise L.
? REGISTRATION NUMBER: 33,732
? REFERENCE/DOCKET NUMBER: UTSD:401
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 713-787-1400
? TELEFAX: 713-789-2679
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 5173 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: CDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..4863
?
US-08-242-677-1

```

```

Query Match Similarity      2.7%; Score 70.4; DB 1; Length 5173;
Best Local Similarity      47.0%; Pred. No. 7,4e-06;
Matches 218; Conservative  0; Mismatches 246; Indels  0; Gaps  0;

Db      105 GAGGCGGTGSCGGTGTGCCCGCGCGCGGTGGGGTCTTCTTCTCTGTGCGCGGGGCGCGGGGCGCG 164
Db      469 GGGCCGCTACTGAGAGCGGTGTGGCGGGAGCCCGCCGTGCGCTTGTGCGCTTGTGGCGGGGCGGG 528
Db      165 GCAGGCGACGAGGCGCCCGGAGGCGCGCGCGCGCGCTGTGTGCGCGCGCTGTGTGCGCGCGCTGTGGGG 224
Db      529 GACGGGAGTAGAGCGCGGCGCTGTGCGAGGACCGCGCGCGCGCGCTGTGTGTGGCGCGGCGGACTGTG 588
Db      225 CCAGGCGCCCGCGCGCGCGACTTCTCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 284
Db      589 CCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 648
Db      285 TTGTGACACTTACAGCTGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 344
Db      649 CCTGGGGGCGTCTGTGGGGTCCGGCGCGCGCTTAGAGGAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 708
Db      345 GAGCTGTGCGCGCGCGCGCGCGCGCTGTGACCGCTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 404
Db      769 GCCGAGAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 768
Db      405 GCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 464
Db      769 CCGGAGCGCGCGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 828
Db      465 ACCGAGCGCACGCGCCCAACAGGTACCGCTATTACCAAAATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 524
Db      829 GACGCGCTGTGACCGCGGACCGAGCGCGGCTTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 888
Db      525 TTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 568
Db      889 GAGCTGTGGGGCGGACTGTGACCTGTGCGGGCGCCGAGGAAAGAAACGG 932

RESULT 4
US-08-440-856A-9
; Sequence 9, Application US/08440856A
; Patent No. 5750873
; General Information:
; APPLICANT: DELAPORTA, STEPHEN L.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR PRODUCING
; TITLE OF INVENTION: PLANTS WITH SINGLE-SEX FLOWERS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
;

```

ADDRESSEE: MORRISON & FORSTER  
STREET: 2000 PENNSYLVANIA AVE. N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/440,856A  
FILING DATE: 15-MAY-1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: MILLMAN, ROBERT A.  
REGISTRATION NUMBER: 36,217  
REFERENCE/DOCKET NUMBER: 05463-20001.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1517  
TELEFAX: (202) 887-0763  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1288 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-440-856A-9

	Query March	2.6%; Score 66.2; DB 1; Length 1288;	
	Best Local Similarity	42.5%; Pred No. 4.1e-05;	
	Matches 155; Conservative 31; Mismatches 179; Indels 0; Gaps 0		
Qy	114 GCAGTGGCCGCGCGCGGTGGGGGTCCTTCTCCTTGCGCCGGGGCCGGGGGCCGCGACGAC	173	
Dp	360 GCGCGGGSGAAGCCGCTGGCGKCGCGCGCTGGGSSCCGCASGTCHRSTTCGTGCGSTGCAC	419	
Qy	174 GAAGCCCCGGAGCGCGCGCCCTGGGGGCGCTCTGTGACC CGAGCTGTGGGCCCAAGCCCC	233	
Dp	420 GTGTGCGTAGAGASGACTSRSRGGCGCCCGTSAGCTGSSGGGCTGTSGCCSRVGCSS	479	
Qy	234 GCGGCGGACTTCTCCGCTGTCGCTGGAGGCGCGCTTGCGTGC CAAGCGGGCTTTGAACCC	293	
Dp	480 CCGCTGAACTSTWSYGCACAACAACCGCGGGTGCTGGGCGCCAGACCGCGCGCCAG	539	
Qy	294 TACAGCTTGCGCGCGCGCGCGCGCGCGCGCTGTCGCGGTGCGCGCTTCAACGCGCGTGGCG	353	
Dp	540 AGCATCTGTCTTCGACGCGSSAGATTGACGCGCGTGTCCGCGTCAACGCGCTGGGG	599	
Qy	354 GCCGCGCGGGGCTGCACCGCTAACCTTGGCGCACTTCTGTGCTGTCCAGTGGCTTGATCC	413	
Dp	600 GCCGCGCTCGGSATGAGACACCGCGCGCKCGCATGACCCAGCCGCCCGCGCGCGGAGA	659	
Qy	414 GGCTCTGAGTGGCGCGCTGCGCGGCGCACGTCAGCGCGCGCGGGGAGCTGACGAGGCC	473	
Dp	660 GCATCTCTCTCGTGCACAGCGTGGSGSCGTGCTBSGGCGGCTTGAGCCGCCACCTTA	719	
Qy	474 ACGCC 478		
Dp	720 CCGCC 724		

RESULT 5  
US-09-128-155-16/c  
; Sequence 16, Application US/09128155  
; Patent No. 6117654  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Yang  
; TITLE OF INVENTION: NOVEL MOLECULES OF TANCO-77 RELATED PROTEIN FAMILY  
; TITLE OF INVENTION: AND USES THEREOF

```

? FILE REFERENCE: 09404/052001
? CURRENT APPLICATION NUMBER: US/09/128,155
? CURRENT FILING DATE: 1998-08-03
? EARLIER APPLICATION NUMBER: US 60/091,650
? EARLIER FILING DATE: 1998-07-02
? EARLIER APPLICATION NUMBER: US 60/054,646
? EARLIER FILING DATE: 1997-08-04
? NUMBER OF SEQ ID NOS: 18
? SOFTWARE: FastSeq For Windows Version 3.0
? SEQ ID NO 16
? LENGTH: 152331
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc feature
? LOCATION: (1)..(152331)
? OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

```

Query Match	2.6%	Score 66.2	DB 3	Length 15233
Best Local Similarity	52.7%	Pred. No. 0.00012		
Matches 166	Conservative	0	Mismatches 148	Indels 1
<p>           Db 72 GGCCTGCGCGGAACCCGACAGACTGATGACCATGAGAGCGGTGCGGTGCGCGCGGTG 131            22261 GGGGGGTGTGGGGGGGGCGCGCGGTGCGGGGTGTGGGGGGGGGGGGGGCGCGCGCG 22202            132 GGGGTCTCTTCTCTGTGGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 191            22201 GGGGGGGGGGGCGCGGG 22144            192 GCCGTGCGAGCGCTGTGTCGCCCGCGCTGTGGAGGACAGGCCCGCGGCGCACTTCTCCGTG 251            22142 GCGGGGCGCGGCGGG 22083            252 TCGGTGGAGCGCGCTCTGTGTGTCGCAACCGCGGCTTGGACACTGACAGCGCTGTGGCGGG 311            22082 GGGGGTCCGTGTGCGCGGGGTGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 22022            312 GCGCGCGCGCGCGCTGTGCGGGTGTGCGGCTCCACAGGCGTGTGCGCGCGCGCGGGCTGCA 371            22022 GGGGTGGGGGGGGCGCGGG 21963            372 CGCTACTGTGCGGAC 386            Db 21962 CGCGCGGGCGCGGCGC 21948         </p>				

Matches 187; Conservative 0; Mismatches 188; Indels 2; Gaps 1;

QY 35 GCCGCGACCTATTGAGACGGGGGCGCCCAACCCCTGCGCGGAGACCCCAAGAC 94  
 Db 3933833 GCGGCGAACCGGCGGACCGGCGGCAAGCGGCGCGTGTGTGGGAACTTGGGGCTC 3933832  
 QY 95 TGAACCATGAGAGCGGTGGCGGTGGCGCGCGGTGGGGGCTCTTCTCTGCGCGGAGC 154  
 Db 3933893 GCGGCGCTCGGCGGACCGGCGGCGGAGACCGCGCATGCGGCTGTGTGTGGCGCGG 3933952  
 QY 155 CCGGGGCGCGGACGCGGACGAGAGCCCGGAGCGCGCGCGCTGGGGGCGCTGAGGCGCG 214  
 Db 3933953 TTGCGCGGGGCGCGGAGTAAAGCCCGCGCGCGGAAACGCGCGCGCTGCGGTGAAGC 3934012  
 QY 215 GCTGCTGGGGCCAGGACCCCGCGCGGACTTCTCTGCTGTGCGGTGAGCGCGCTGAGCTG 274  
 Db 3934013 GCGCATGGGGCCAGAGGCTCTGCGCTGCGGCTCTGCGGCTTTGAACGCGCGCAAGCGCG 3934072  
 QY 275 CAAGCGCGG--CTTGAACACTTACAGCTTGGGCGCGCGCGCGCGCGCGCGGTG 332  
 Db 3934073 CAAGCGCGGCGCGCGCGGACCGCGCGCGCGCGCATCAACGGGCGCGCGCGCGCG 3934132  
 QY 333 CCGGCGCTCCACCGGGGCGTGGCGCGCGCGCGCGGAGGCTGACCGCTACTGCGGACTTGT 392  
 Db 3934133 GCGCAACGCGCGGCGGACCGGCGGCGGCGGCGCAACCGGTGCGGAGGTCTCGCGAACAGCG 3934192  
 QY 393 GCGTGCACAGTGGCGCTG 409  
 Db 3934193 GGGGTGCGGCGGTGACGG 3934209

## RESULT 7

US-09-103-840A-2  
 ; Sequence 2, Application US/09103840A  
 ; Patent No. 6294328

; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; APPLICANT: WHITE, Owen R.  
 ; APPLICANT: FRASER, Claire M.  
 ; APPLICANT: VENTER, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; TITLE OF INVENTION: TUBERCULOSIS  
 ; FILE REFERENCE: 24366-20007.00  
 ; CURRENT APPLICATION NUMBER: US/09/103,840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 4403765  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; FEATURE:  
 ; OTHER INFORMATION: CDC 1551  
 ; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
 ; OTHER INFORMATION: represent a, t, c or g  
 ; US-09-103-840A-2

Query Match 2.4%; Score 61; DB 3; Length 4403765;  
 Best Local Similarity 49.1%; Pred. No. 0.0031;

Matches 189; Conservative 0; Mismatches 195; Indels 1; Gaps 1;

QY 94 CTGAGACCATGAGAGCGGTGGCGGTGGCGCGCGGTGGGGTCTCTTCTCTGCGCGG 153  
 Db 625458 CTTGCGGACAGGCGGAGCGGTGCGCGGTGTATGAGGCTGTGCGTGGGCGCGGTGGCGG 625517  
 QY 154 CCGGGGCGCGGACGCGGACGAGGCGCGGAGCGCGCGCGCTGCGGCGCTGCTGCGCC 213  
 Db 625518 CCGGTGCTTGGGACAGGTGGCGCGCGCGCGGTGCGCGCGCGCGCGGACACAC 625577  
 QY 214 GGTGCTGGGCGGAGCGCGCGCGGCGGACTTCTCTGCTGTGCGGTGAGCGCGCTGCGCTG 273  
 Db 625578 CGGCATCGGTGGGACAGGCGGCGGTGCGGTGGGCGCGGTGGGTGGGTGATACG 625637

QY 274 CCAAGCGGCGCTTGAACACTTACAGCTTGGCGCGCGCGCGCGCGCGCGCGCTGCGGCTG 333

Db 625638 CCAAGCGGCGCGCGCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 625697

QY 334 GCGGCGTCCACGGGCGGTGGCG 393

Db 625698 TGGCGACGCGGCGGAGACG 625756

QY 394 GCTGCGACGTGGCGCTTGTGCGCGCTTCACTGCGCGCTTGTGCGCGCGCGCGCGCGCG 453

Db 625757 ACAGCGGCG 625816

QY 454 CCGGCGGACGTGACCGGACGCGCGCG 478

Db 625817 GCGCGGCGGCGCGCGCGCGCGCGCGCG 625841

## RESULT 8

US-09-103-840A-2/c  
 ; Sequence 2, Application US/09103840A  
 ; Patent No. 6294328

; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; APPLICANT: WHITE, Owen R.  
 ; APPLICANT: FRASER, Claire M.  
 ; APPLICANT: VENTER, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; TITLE OF INVENTION: TUBERCULOSIS  
 ; FILE REFERENCE: 24366-20007.00  
 ; CURRENT APPLICATION NUMBER: US/09/103,840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 4403765  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; FEATURE:  
 ; OTHER INFORMATION: CDC 1551  
 ; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
 ; OTHER INFORMATION: represent a, t, c or g  
 ; US-09-103-840A-2

Query Match 2.3%; Score 60.2; DB 3; Length 4403765;  
 Best Local Similarity 54.8%; Pred. No. 0.0045;

Matches 143; Conservative 0; Mismatches 113; Indels 5; Gaps 1;

QY 109 CCGTGGCGGTGGCGCGCGCGCGCGCGGTCTTCTCTGCGCGCGCGCGCGCGCGCGCGCGAG 168  
 Db 337256 CCGGCGGCG 337197  
 QY 169 GCGACGAGCGCGGAGCG 228  
 Db 337196 GCTCTTGGCG 337137  
 QY 229 GCCCG 283  
 Db 337136 GGGCGACG 337077  
 QY 284 CTTGACACCTTACAGCTTGGCG 343  
 Db 337076 CACTTGGCG 337017  
 QY 344 GGGCGTGGCGCGCGCGCGCGCGCG 364  
 Db 337016 CCGCGCGCGCGCGCGCGCGCGCGCGCGCG 336996

## RESULT 9

US-09-103-840A-1/c  
 ; Sequence 1, Application US/09103840A  
 ; Patent No. 6294328

; GENERAL INFORMATION:



QY 27 GCGCGAGGCGCGGACGCTGATTGAGACGCGGCGCGGCCACCCCTCGGCGCGGAGC 86  
 Db 2045 GCGCGGTCGCGCGCTCAGAGTGCCTGCGCGCGCGGATCCTGAGAGCGCTGCGCGAGG 2104  
 QY 87 GCGAGACTGAGACCATGAGAGCGGCTGAGCGCGCGCGGATCTTCTCTCTG 146  
 Db 2105 GCTTGAACGCGGACCT-TGGCGGCGCGTCCCGGCGCGCGCGCGCGCGCGCGCGCG 2163  
 QY 147 GCGCGGCGCGGCGCGCGCGCGCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCTC 206  
 Db 2164 CCGCGCGCGCGAGAGACCCCGCGCGCGCGCGCGCTTCCCGCGCGCGCGCGCGAGCGCGCG 2223  
 QY 207 GTGGCGCGCGCTGCTGAGGCGCGAGCGCGCGCGCGCGCGCTTCTCGGTCGAGGCGCGCT 266  
 Db 2224 CGCGTCGCGCGCTGAGTGTGCGGACGCTGCGGCTTCTGCGCGAGCGCGCGCGCGCG 2283  
 QY 267 CTGGCTGCGAGCGCGGCTTGAACACTACAGCTGCGCGCGCGCGCGCGCGCGCGCTG 326  
 Db 2284 CTGCGCGGAGGACCTGCGCGCTGCGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCG 2343  
 QY 327 CGGTCGCGGCGCTTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 386  
 Db 2344 GTAGCTGTGTGCG 2403  
 QY 387 TTCTGTGCTGCGCACGCTGCGCTGTCTCGGCTCTCACTGCGCGCGCGCGCGCGCGCA 446  
 Db 2404 TCCGCGCGCGCGCGCGCGCGCGCGAGACTGTGTTGACACAGAGCGCTGCGCGCGCG 2463  
 QY 447 GCGGTGCGCG 456  
 Db 2464 GCGCGCGCGG 2473

# RESULT 12

US-08-843-659-1  
 / Sequence 1, Application US/08843659  
 / Patent No. 6218103  
 / GENERAL INFORMATION:  
 / APPLICANT: Leopardi, Rosario  
 / APPLICANT: Roizman, Bernard  
 / TITLE OF INVENTION: HERPES SIMPLEX VIRUS US3 AND ICP4 AS  
 / TITLE OF INVENTION: INHIBITORS OF APOPTOSIS  
 / NUMBER OF SEQUENCES: 6  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Arnold, White & Durkee  
 / STREET: P.O. Box 4433  
 / CITY: Houston  
 / STATE: Texas  
 / COUNTRY: United States  
 / ZIP: 77210  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: Patent in Release #1.0, Version #1.30  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/843,659  
 / FILING DATE: Concurrently Herewith  
 / CLASSIFICATION: 435  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Highlander, Steven L.  
 / REGISTRATION NUMBER: 37,642  
 / REFERENCE/DOCKET NUMBER: ARS:519  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (512) 418-3000  
 / TELEFAX: (512) 474-7577  
 / INFORMATION FOR SEQ ID NO: 1:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 4257 base pairs  
 / TYPE: nucleic acid  
 / STRANDEDNESS: single  
 / TOPOLOGY: linear  
 / US-08-843-659-1

Query Match 2.3%; Score 59.6; DB 3; Length 4257;  
 Best Local Similarity 47.7%; Pred. No. 0.0013;  
 Matches 205; Conservative 0; Mismatches 224; Indels 1; Gaps 1;

QY 27 GCGCGAGGCGCGGACGCTGATTGAGACGCGGCGCGGCCACCCCTCGGCGCGGAGC 86  
 Db 2045 GCGCGGTCGCGCGCTCAGAGTGCCTGCGCGCGCGGATCCTGAGAGCGCTGCGCGAGG 2104  
 QY 87 GCGAGACTGAGACCATGAGAGCGGCTGAGCGCGCGGATCTTCTCTCTG 146  
 Db 2105 GCTTGAACGCGGACCT-TGGCGGCGCGTCCCGGCGCGCGCGCGCGCGCGCGCGCG 2163  
 QY 147 GCGCGGCGCGGCGCGCGCGCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCTC 206  
 Db 2164 CCGCGCGCGAGAGACCCCGCGCGCGCGCGCGCTTCCCGCGCGCGCGCGAGCGCGCG 2223  
 QY 207 GTGGCGCGCGCTGCTGAGGCGCGAGCGCGCGCGCGCGCGCTTCTCGGTCGAGGCGCGCT 266  
 Db 2224 CGCGTCGCGCGCTGAGTGTGCGGACGCTGCGGCTTCTGCGCGAGCGCGCGCGCGCG 2283  
 QY 267 CTGGCTGCGAGCGCGGCTTGAACACTACAGCTGCGCGCGCGCGCGCGCGCGCGCTG 326  
 Db 2284 CTGCGCGGAGGACCTGCGCGCTGCGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCG 2343  
 QY 327 CGGTCGCGGCGCTTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 386  
 Db 2344 GTAGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGAGACTGTGTTGACACAGAGCG 2403  
 QY 387 TTCTGTGCTGCGCACGCTGCGCTGTCTCGGCTCTCACTGCGCGCGCGCGCGCGCGCA 446  
 Db 2404 TCCGCGCGCGCGCGCGCGCGCGCGAGACTGTGTTGACACAGAGCGCTGCGCGCGCG 2463  
 QY 447 GCGGTGCGCG 456  
 Db 2464 GCGCGCGCGG 2473

# RESULT 13

US-08-458-568A-11/C  
 / Sequence 11, Application US/08458568A  
 / Patent No. 5821339  
 / GENERAL INFORMATION:  
 / APPLICANT: Schaffer, Priscilla A.  
 / APPLICANT: Yeh, Lily  
 / TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus  
 / TITLE OF INVENTION: Infections  
 / NUMBER OF SEQUENCES: 15  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339-15  
 / STREET: One Liberty Place, 46th floor  
 / CITY: Philadelphia  
 / STATE: PA  
 / COUNTRY: USA  
 / ZIP: 19103  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: Wordperfect 5.1  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/458,568A  
 / FILING DATE: 02-JUNE-1995  
 / CLASSIFICATION: 435  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 08/065,146  
 / FILING DATE: 05-MAY-1993  
 / CLASSIFICATION: 435  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Leary Ph.D., Kathryn R.  
 / REGISTRATION NUMBER: 36,317  
 / REFERENCE/DOCKET NUMBER: DFCT-0029  
 / TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12001 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Herpes simplex virus  
STRAIN: Herpes Simplex Virus Type 1  
US-08-458-568A-11

Query Match 2.3% Score 59.6; DB 1; Length 12001;  
Best Local Similarity 47.7%; Pred. No. 0.0015;  
Matches 205; Conservative 0; Mismatches 224; Indels 1; Gaps 1;

QY 27 GCGCGAGGCGCGGAGCTGATTGACGCGGCGCGCCCGCCCGCTGCGCGCGGAGCC 86  
Db 4629 GCGCGAGGCGCGGAGCTGATTGACGCGGCGCGCCCGCCCGCTGCGCGCGGAGCC 4570  
QY 87 GCGAGGAGTGAACCATGAGAGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 146  
Db 4569 GCTTCGAGCGGAGCC-TGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4511  
QY 147 GCGCGGCG 206  
Db 4510 CCGCGGCGCGGAGGAGCG 4451  
QY 207 GTGCG 266  
Db 4450 CGCGCTGCG 4391  
QY 267 CTGCGCTCCAGCG 326  
Db 4390 CTGCGCGCGGAGCG 4331  
QY 327 CGGCTGCG 386  
Db 4330 GTGAGCGCTGCG 4271  
QY 387 TTCTGTGCTGCG 446  
Db 4270 TCG 4211  
QY 447 GCGCGTGCAG 456  
Db 4210 GCGCGCGCGCG 4201

RESULT 14  
US-08-118-200-1  
Sequence 1, Application US/08118200  
Patent No. 6197500

GENERAL INFORMATION:  
APPLICANT: SUTHERLAND, Grant R  
APPLICANT: RICHARDS, Robert I  
APPLICANT: SCHLESINGER, David  
APPLICANT: NAGARAJA, Ramalah  
APPLICANT: KREMER, Eric J  
APPLICANT: YU, Sui  
APPLICANT: BAKER, Elizabeth  
APPLICANT: MULLEY, John C  
APPLICANT: MANDEL, Jean-Louis  
APPLICANT: PRITCHARD, Melanie April  
APPLICANT: LINCX, Michael  
TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
City: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/118,200  
Filing DATE: 09-SEP-1993  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/802,650  
Filing DATE: 05-DEC-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,232  
Filing DATE: 20-MAR-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/638,518  
Filing DATE: 04-JAN-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/966,517  
Filing DATE: 23-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 020160-164  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1028 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-118-200-1

Query Match 2.3% Score 58.6; DB 3; Length 1028;  
Best Local Similarity 59.2%; Pred. No. 0.0015;  
Matches 100; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 70 CTGCGCGTTCGCGGAGCCCGCAGAGCTGAGACCATGAGGCGGCTGCGCGCGCGCG 129  
Db 340 CCGGCGGAGCGGAGCG 399  
QY 130 TGGGGGTCCTTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 189  
Db 400 CCGCGCGGCTGCG 459  
QY 190 CCGCGCGTGCAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 238  
Db 460 CCGCGCGGCG 508

RESULT 15  
US-08-458-745-1  
Sequence 1, Application US/08458745  
Patent No. 6242576

GENERAL INFORMATION:  
APPLICANT: SUTHERLAND, Grant R  
APPLICANT: RICHARDS, Robert I  
APPLICANT: SCHLESINGER, David  
APPLICANT: NAGARAJA, Ramalah  
APPLICANT: KREMER, Eric J  
APPLICANT: YU, Sui  
APPLICANT: BAKER, Elizabeth  
APPLICANT: MULLEY, John C

APPLICANT: MANDEL, Jean-Louis  
APPLICANT: PITCHARD, Melanie April  
APPLICANT: LYNCH, Michael  
TITLE OF INVENTION: DNA SEQUENCES RELATED TO ISOLATED  
TITLE OF INVENTION: FRAGILE X SYNDROME  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,745  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/118,200  
FILING DATE: 09-SEP-1993  
APPLICATION NUMBER: US 07/802,650  
FILING DATE: 05-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,232  
FILING DATE: 20-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/638,518  
FILING DATE: 04-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/966,517  
FILING DATE: 23-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feuyl, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 020160-164  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-5620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1028 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-458-745-1

Query Match 2.3%; Score 58.6; DB 3; Length 1028;  
Best Local Similarity 59.2%; Pred. No. 0.0015;  
Matches 100; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 70 CTGGCGGTGGCGGAGCCCGGAGCATGAGAGCGGTGGCGCGCGCGG 129  
DB 340 CGGCGCGACGGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCG 399  
QY 130 TGGGGGTCTTCTCTGCGCGGCGCGCGCGCGCGCGCGCGCGAG 189  
DB 400 CGGCGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 459  
QY 190 CGGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 238  
DB 460 CGGCG 508

Search completed: February 13, 2004, 13:34:51  
Job time : 179 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 13, 2004, 12:03:44 ; Search time 890 Seconds  
(without alignments)  
10657.701 Million cell updates/sec

Title: US-09-836-613-1

Perfect score: 2575  
Sequence: 1 CCCGGGCTTAGACCTTCGGGT.....AGTCAGCGCGCCGAATTC 2575

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCr\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCrUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2575	100.0	2575	11	US-09-836-613-1
2	1426.6	55.4	10380	11	US-09-836-613-3
3	1211	47.0	1211	13	US-10-029-386-22784
4	485.6	18.9	2796	13	US-10-027-632-111960
5	485.6	18.9	2796	13	US-10-027-632-111960
6	456.6	17.7	808	14	US-10-027-632-141597
7	456.6	17.7	808	14	US-10-027-632-141597
8	445.4	17.3	499	11	US-09-918-995-30489
9	426	16.5	522	13	US-10-029-386-9071
10	395	15.3	449	11	US-09-918-995-24382
11	358.4	13.9	468	11	US-09-918-995-21830
12	183.4	7.1	2184	15	US-10-156-761-2002
13	183.4	7.1	2184	15	US-10-156-761-2002
14	155	6.0	3114	15	US-10-156-761-5965
15	155	6.0	9025608	15	US-10-156-761-1

16	88	3.4	2202	12	US-10-260-238-900	Sequence 900, App
17	73.6	2.9	11817	15	US-10-156-761-2884	Sequence 2884, Ap
18	73.6	2.9	125746	15	US-10-156-761-15102	Sequence 15102, A
19	72.6	2.8	1011	13	US-10-029-386-25041	Sequence 25041, A
20	71	2.8	4020	11	US-09-796-679-5	Sequence 5, Appli
21	70.4	2.7	5173	15	US-10-171-561-159	Sequence 3356, Ap
22	70.4	2.7	5173	15	US-10-171-561-159	Sequence 159, App
23	70	2.7	10144	10	US-09-880-107-2168	Sequence 2168, Ap
24	69.2	2.7	51657	12	US-10-057-475B-10475	Sequence 10475, A
25	69.2	2.7	51657	12	US-10-154-884B-10475	Sequence 48, Appl
26	68.6	2.7	2561	10	US-09-976-740-48	Sequence 48, Appl
27	68.6	2.7	2561	12	US-10-616-187-48	Sequence 48, Appl
28	68.6	2.7	2561	14	US-10-023-523-48	Sequence 48, Appl
29	68.6	2.7	2561	14	US-10-023-523-48	Sequence 48, Appl
30	68.2	2.6	2271	13	US-10-101-510-133	Sequence 133, App
31	68.2	2.6	2487	15	US-10-205-823-135	Sequence 15, App
32	66.2	2.6	152331	14	US-10-095-407-16	Sequence 16, Appl
33	64.4	2.5	1614	10	US-09-976-740-45	Sequence 45, Appl
34	64.4	2.5	1614	12	US-10-616-187-45	Sequence 45, Appl
35	64.4	2.5	1614	14	US-10-023-523-45	Sequence 45, Appl
36	64.4	2.5	1614	14	US-10-023-523-45	Sequence 45, Appl
37	64.4	2.5	3957	13	US-10-200-562-193	Sequence 193, App
38	64.4	2.5	3957	13	US-10-237-551-193	Sequence 193, App
39	64.4	2.5	4225	13	US-10-141-541-3	Sequence 3, Appli
40	64.4	2.5	12425	10	US-09-976-740-50	Sequence 50, Appl
41	64.4	2.5	12425	12	US-10-616-187-50	Sequence 50, Appl
42	64.4	2.5	12425	14	US-10-023-523-50	Sequence 50, Appl
43	64.4	2.5	154746	13	US-10-023-523-50	Sequence 50, Appl
44	64.4	2.5	154746	13	US-09-827-668-8	Sequence 8, Appli
45	64.4	2.5	154746	13	US-09-827-668-8	Sequence 8, Appli

#### ALIGNMENTS

RESULT 1  
US-09-836-613-1  
Sequence 1, Application US/09836613  
Publication No. US20030039643A1  
GENERAL INFORMATION:  
APPLICANT: HOWARD, JOHN JOSEPH; SCOTT, HAMISH STEPHEN;  
TITLE OF INVENTION: SYNTHETIC MAMMALIAN  
N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES  
ENCODING SAME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSER: NIXON PEABODY LLP  
STREET: 990 STEWART AVENUE  
CITY: GARDEN CITY  
STATE: NEW YORK  
COUNTRY: UNITED STATES  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/836,613  
FILING DATE: 17-Apr-2001  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00747  
FILING DATE: 22-Nov-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: POKALSKY, ANN R.  
REGISTRATION NUMBER: 34,697  
REFERENCE/DOCKET NUMBER: 2249/104  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516 742 4366  
TELEFAX: 516 742 4366  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

LENGTH: 2575 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Peripheral Blood  
CELL TYPE: Leukocyte  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 102..2330  
SEQUENCE DESCRIPTION: SEQ ID NO: 1  
US-09-836-613-1

Query Match 100.0%; Score 2575; DB 11; Length 2575;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CCCGGGCTTACCTTCGGGTCACGTGCGCGAGGCCGAGCTGATTGAGCGCGGGCCG 60
DB 1 CCCGGGCTTACCTTCGGGTCACGTGCGCGAGGCCGAGCTGATTGAGCGCGGGCCG 60
QY 61 CCCGACCCCTGCGCGCGTCGCGAGACCCGAGACTGAGACATGAGAGCGGTCGCTG 120
DB 61 CCCGACCCCTGCGCGCGTCGCGAGACCCGAGACTGAGACATGAGAGCGGTCGCTG 120
QY 121 CCGCGGCGGTGAGGAGTCTTCTCCTGAGCCGAGGCGCGGAGCGGAGCGAGCGAG 180
DB 121 CCGCGGCGGTGAGGAGTCTTCTCCTGAGCCGAGGCGCGGAGCGGAGCGAGCGAG 180
QY 181 GGGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
DB 181 GGGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 241 ACTTCTCGTGTGCGGTGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
DB 241 ACTTCTCGTGTGCGGTGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
QY 301 TGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
DB 301 TGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
QY 361 CGGGGCTGCAACCGTACTGCGGACCTTCTGTGCTGCGACGTCGCTGCTGCTGCT 420
DB 361 CGGGGCTGCAACCGTACTGCGGACCTTCTGTGCTGCGACGTCGCTGCTGCTGCT 420
QY 421 AGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
DB 421 AGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
QY 481 ACAGGTACCGGTATTACAGATGTGTGCAAGCAAGCTACTCTTCTGCTGCTGCTG 540
DB 481 ACAGGTACCGGTATTACAGATGTGTGCAAGCAAGCTACTCTTCTGCTGCTGCTG 540
QY 541 GGGCGCGCTGCGAGCGAGATAGACTGATGCGCTGATGCGATCAACTGCGACTG 600
DB 541 GGGCGCGCTGCGAGCGAGATAGACTGATGCGCTGATGCGATCAACTGCGACTG 600
QY 601 CTTGAGGCGCGCGAGAGCGCATCTGCGAGCGGAGTACCTGCGCTTGGGCTGACCA 660
DB 601 CTTGAGGCGCGCGAGAGCGCATCTGCGAGCGGAGTACCTGCGCTTGGGCTGACCA 660
QY 661 CAGAGATCAATGATTTCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
DB 661 CAGAGATCAATGATTTCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 721 TGGACACTGCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
DB 721 TGGACACTGCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
QY 781 ACCGGGCTCTGAGCAAGATGCGCTCTTCCGAGTGAACCCAGTGTGCTGCTGCTG 840
```

```
DB 781 ACCGGGCTCTGAGCAAGATGCGCTCTTCCGAGTGAACCCAGTGTGCTGCTGCTG 840
QY 841 GGCATGTTCCCGAGGCTGTACACAGGGGTTCCTCAGGTCAATGTACAGAAATGGACA 900
DB 841 GGCATGTTCCCGAGGCTGTACACAGGGGTTCCTCAGGTCAATGTACAGAAATGGACA 900
QY 901 GTTGGGCGCACTTTAATCTTCTTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
DB 901 GTTGGGCGCACTTTAATCTTCTTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
QY 961 TATTCCCATATGAGGAGGCTCTTCTGCGAGAGCGTATTAAGAGTTGGCAGAGCC 1020
DB 961 TATTCCCATATGAGGAGGCTCTTCTGCGAGAGCGTATTAAGAGTTGGCAGAGCC 1020
QY 1021 ACATCTATGAGGCGGACACTTTCAATGATGATGAGCCACCTTCTCAGAGCCCTCT 1080
DB 1021 ACATCTATGAGGCGGACACTTTCAATGATGATGAGCCACCTTCTCAGAGCCCTCT 1080
QY 1081 TTGCGCGAGCCACACATGCGCTCTATGAGGCGCATGACTGAGTGAATGAGGCTGCT 1140
DB 1081 TTGCGCGAGCCACACATGCGCTCTATGAGGCGCATGACTGAGTGAATGAGGCTGCT 1140
QY 1141 GGCCTGCTCAAGGCTGCTCTTCCAGACACAGCCGAGTTCTGAGGCGCCGCCAGATCA 1200
DB 1141 GGCCTGCTCAAGGCTGCTCTTCCAGACACAGCCGAGTTCTGAGGCGCCGCCAGATCA 1200
QY 1201 GGGCTGTCTGAGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
DB 1201 GGGCTGTCTGAGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
QY 1261 GCGAGCTGTATATACCGGCACTGCTCTTCCAGAGGCGAGCCCTCATCTGTGTCATG 1320
DB 1261 GCGAGCTGTGTATATACCGGCACTGCTCTTCCAGAGGCGAGCCCTCATCTGTGTCATG 1320
QY 1321 TGCACACTTTGGGGGAAACCATGCTTTTGGAGCCCTTATGAGGCTGTGAGCGAGGCC 1380
DB 1321 TGCACACTTTGGGGGAAACCATGCTTTTGGAGCCCTTATGAGGCTGTGAGCGAGGCC 1380
QY 1381 CAGAGCTGCGCGCGCTCTTCCCAACTCCACATGATGAGGAGCGGCGATGCGCGAGG 1440
DB 1381 CAGAGCTGCGCGCGCTCTTCCCAACTCCACATGATGAGGAGCGGCGATGCGCGAGG 1440
QY 1441 GCATAGCCAGAGAGATGATGCTATTTCCCTCATGAGCTGAGCTGAGCGGAGAAAGACC 1500
DB 1441 GCATAGCCAGAGAGATGATGCTATTTCCCTCATGAGCTGAGCTGAGCGGAGAAAGACC 1500
QY 1501 CAGTCCCAATTTTGGCAAGCTTGGGTGACCAAGTTTGCCTGCGCGCGGTATGAGGCT 1560
DB 1501 CAGTCCCAATTTTGGCAAGCTTGGGTGACCAAGTTTGCCTGCGCGCGGTATGAGGCT 1560
QY 1561 ACCCGGAGCGAGGCGAGCGTGAAGGCTACTGCTCCGAGAGTGTGTACAAGTCTCCG 1620
DB 1561 ACCCGGAGCGAGGCGAGCGTGAAGGCTACTGCTCCGAGAGTGTGTACAAGTCTCCG 1620
QY 1621 AGGCTTGAAGGCGGCAATGTATGCGCGCTGTCAGAGCGGCGCTCTTACATGATGATA 1680
DB 1621 AGGCTTGAAGGCGGCAATGTATGCGCGCTGTCAGAGCGGCGCTCTTACATGATGATA 1680
QY 1681 CAGGATCTGTGTACAACGATCTGATGTGTGTGAGGCTGAGGCGCTGCTGCTCATCTG 1740
DB 1681 CAGGATCTGTGTACAACGATCTGATGTGTGTGAGGCTGAGGCGCTGCTGCTCATCTG 1740
QY 1741 CTCCTCTCTTGGCAACAGCCCGCGCTTCCGCTACAGACTGCTGCTGCTGCTGCTG 1800
DB 1741 CTCCTCTCTTGGCAACAGCCCGCGCTTCCGCTACAGACTGCTGCTGCTGCTGCTGCTG 1800
QY 1801 CAGTGAAGAGCTGATGAGTGTACTATGAGAGGCAAGAGCGCTTACTGTGAGCAAG 1860
DB 1801 CAGTGAAGAGCTGATGAGTGTACTATGAGAGGCAAGAGCGCTTACTGTGAGCAAG 1860
QY 1861 AGTGGCTCTCTTGTAGAGGCTGAGAGCGTCTGCGCTATGAGCTGCTGCGGCACTGG 1920
DB 1861 AGTGGCTCTCTTGTAGAGGCTGAGAGCGTCTGCGCTATGAGCTGCTGCGGCACTGG 1920
```

QY 1921 ACAGAGTGTGCTAGTGAACAGCGCTTCTTGTGAGGCACTGAGTGAAGAGCCGAG 1980  
 DB 1921 ACAGAGTGTGCTAGTGAACAGCGCTTCTTGTGAGGCACTGAGTGAAGAGCCGAG 1980  
 QY 1981 CAGGCGCAGTGAAGGCGGAGGCGGATTTCTAGAGCAAGAGCGGCTTACCACTGA 2040  
 DB 1981 CAGGCGCAGTGAAGGCGGAGGCGGATTTCTAGAGCAAGAGCGGCTTACCACTGA 2040  
 QY 2041 CTTGTGGGGGCGAAGAGCAATCTGAGCTTGTGCAACAGAGCACTGAGGCGGCTTGA 2100  
 DB 2041 CTTGTGGGGGCGAAGAGCAATCTGAGCTTGTGCAACAGAGCACTGAGGCGGCTTGA 2100  
 QY 2101 TGAGCAACTACTACACCCCTGCTGAGGCGCTTTTCTGAGAGCGCTGTTGAAGTGTG 2160  
 DB 2101 TGAGCAACTACTACACCCCTGCTGAGGCGCTTTTCTGAGAGCGCTGTTGAAGTGTG 2160  
 QY 2161 CCGAGGCGATCCCTTCCAAAGAGAGAGTTCACAAATATGCTTCCAACTGAGAGAG 2220  
 DB 2161 CCGAGGCGATCCCTTCCAAAGAGAGAGTTCACAAATATGCTTCCAACTGAGAGAG 2220  
 QY 2221 CTTTCTGCTAGAGAGAGAGTACCCAGCCAGCCGAGAGAGACACTGTGAGCCTG 2280  
 DB 2221 CTTTCTGCTAGAGAGAGAGTACCCAGCCAGCCGAGAGAGACACTGTGAGCCTG 2280  
 QY 2281 CCAAGAGATTTCTCTCAATATTAACCCCGCTGAGGCTGCGCTCTTGTGATGATTC 2340  
 DB 2281 CCAAGAGATTTCTCTCAATATTAACCCCGCTGAGGCTGCGCTCTTGTGATGATTC 2340  
 QY 2341 GCCACACTGAGGCTTGTTCGCTAATTCAGAGGAGATTCAGAGGCGCCAGAGCTGA 2400  
 DB 2341 GCCACACTGAGGCTTGTTCGCTAATTCAGAGGAGATTCAGAGGCGCCAGAGCTGA 2400  
 QY 2401 CAGACATACAGATTAACCCAGGCTGAGAGAGGCGCCAGGCTTGTGAGGCTG 2460  
 DB 2401 CAGACATACAGATTAACCCAGGCTGAGAGAGGCGCCAGGCTTGTGAGGCTG 2460  
 QY 2461 ACCTGGGGGATTTGAGGAGAAATGACCTGCGCTCAGACCAAGGCTGAGGATTA 2520  
 DB 2461 ACCTGGGGGATTTGAGGAGAAATGACCTGCGCTCAGACCAAGGCTGAGGATTA 2520  
 QY 2521 AGTACTGTTTCTTCCACTTAATAAAAAAAAAAAGTCAGAGCGCGCAATTC 2575  
 DB 2521 AGTACTGTTTCTTCCACTTAATAAAAAAAAAAAGTCAGAGCGCGCAATTC 2575

RESULT 2  
 US-09-836-613-3  
 : Sequence 3, Application US/09836613  
 : Publication No. US20030039643A1  
 : GENERAL INFORMATION:  
 : APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;  
 : WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART  
 : TITLE OF INVENTION: SYNTHETIC MAMMALIAN  
 : N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES  
 : ENCODING SAME  
 : NUMBER OF SEQUENCES: 6  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: NIXON PEARODY LLP  
 : STREET: 990 STEWART AVENUE  
 : CITY: GARDEN CITY  
 : STATE: NEW YORK  
 : COUNTRY: UNITED STATES  
 : ZIP: 11530  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patent Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/09/836,613  
 : FILING DATE: 17-Apr-2001  
 : PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/00747  
 FILING DATE: 22-NOV-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: POKALSKY, ANN R.  
 REGISTRATION NUMBER: 34,697  
 REFERENCE/DOCKET NUMBER: 2249/104  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 516 742 4343  
 TELEFAX: 516 742 4366  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 10380 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 POSITION IN GENOME:  
 CHROMOSOME/SEGMENT: Chromosome 17  
 FEATURE:  
 NAME/KEY: exon 1  
 LOCATION: 990..1372  
 FEATURE:  
 NAME/KEY: exon 2  
 LOCATION: 2115..2262  
 FEATURE:  
 NAME/KEY: exon 3  
 LOCATION: 3056..3202  
 FEATURE:  
 NAME/KEY: exon 4  
 LOCATION: 3387..3472  
 FEATURE:  
 NAME/KEY: exon 5  
 LOCATION: 5667..5923  
 NAME/KEY: exon 6  
 LOCATION: 7745..8955  
 SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
 US-09-836-613-3

Query Match 55.4%; Score 1426.6; DB 11, Length 10380;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1429; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1112 CATGACTGAGTGAATGAGAGGCTGTGTGCTGCTCCAGAGCTGCTCTTCCAGACCA 1171  
 DB 7734 CTTGTCCAGTGAATGAGAGGCTGTGTGCTGCTCCAGAGCTGCTCTTCCAGACCA 7793  
 QY 1172 GCGGAGTTCGTGGGGGCGCGCCAGATCAGGCTGTGTGAGGCTGTGCGCG 1231  
 DB 7794 GCGGAGTTCGTGGGGGCGCGCCAGATCAGGCTGTGTGAGGCTGTGCGCG 7853  
 QY 1232 CCTCTGCTTCTGACACTGTTTGTCTGAGAGCAGGCTGTATACCCGACGCTCCT 1291  
 DB 7854 CCTCTGCTTCTGACACTGTTTGTCTGAGAGCAGGCTGTATACCCGACGCTCCT 7913  
 QY 1292 CCAGGCGCAGCCCTCATCTGATGATCTGACCAACTTTGGGGGAAACATGATCTTT 1351  
 DB 7914 CCAGGCGCAGCCCTCATCTGATGATCTGACCAACTTTGGGGGAAACATGATCTTT 7973  
 QY 1352 TGAAGCCTTAAGAGGCTGTGAAGAGAGGCGCCAGAACTGCGCGCTTCCCAATCCAC 1411  
 DB 7974 TGAAGCCTTAAGAGGCTGTGAAGAGAGGCGCCAGAACTGCGCGCTTCCCAATCCAC 8033  
 QY 1412 CATGTAGGCGAGGCGCATGCGCCCGAGAGGCGCATGAGGCGAGAGTGTATTCCT 1471  
 DB 8034 CATGTAGGCGAGGCGCATGCGCCCGAGAGGCGCATGAGGCGAGAGTGTATTCCT 8093  
 QY 1472 CATGCTGAGCTGGGCTGCGGAAAGAGCCAGTGCAGATTTGGAGCTGTGATCCAG 1531  
 DB 8094 CATGCTGAGCTGGGCTGCGGAAAGAGCCAGTGCAGATTTGGAGCTGTGATCCAG 8153

QY	1532	CTTTGGCCGCGCGCGGATATN	GGGGGTCTCCACCGCGAAGCA	GGGGCAAGCGTGAAGCTACT	1597
Db	8154	CTTTGCCGCCGCGCGGATATN	GGGGGTCTCCACCGCGAAGCA	GGGGCAAGCGTGAAGCTACT	8213
QY	1592	GCTCCGAGATGTGTACAACTG	CTCCGGAGAGCCTTCAGAGGCGCA	CAATGTATGCCGCT	1651
Db	8214	GCTCCGAGATGTGTACAACTG	CTCCGGAGAGCCTTCAGAGGCGCA	CAATGTATGCCGCT	8272
QY	1652	GCTCAGGCGCGCGCTCCCTTA	CAATATTAACCAAGATCTGGTA	CAACCGATCTGATGTGT	1711
Db	8274	GCTCAGGCGCGCGCTCCCTTA	CAATATTAACCAAGATCTGGTA	CAACCGATCTGATGTGT	8333
QY	1712	TGAGGCTGTGAGCGCTGTCTCA	CAATGTCTCCCTCCTGAGCA	CAAGCCCGCTTCG	1771
Db	8334	TGAGGCTGTGAGCGCTGTCTCA	CAATGTCTCCCTCCTGAGCA	CAAGCCCGCTTCG	8393
QY	1772	CTACGACTGTGAGACTTCACT	CGGCAAGTGCAGAGACTGTGTCA	CGTGTACTATGA	1831
Db	8394	CTACGACTGTGAGACTTCACT	CGGCAAGTGCAGAGACTGTGTCA	CGTGTACTATGA	8453
QY	1832	GGAAGGCAAGAGAGCGCTAC	CTGAGGCAAGACCTGCGCTCC	CTGTGAGGGCTGAGAGCGT	1897
Db	8454	GGAAGGCAAGAGAGCGCTAC	CTGAGGCAAGACCTGCGCTCC	CTGTGAGGGCTGAGAGCGT	8513
QY	1892	CCTGAGCTTATGAGCTGTCTG	CCGCACTGGAAGAGTCTGAGCA	AGCCGCTTCTT	1951
Db	8514	CCTGAGCTTATGAGCTGTCTG	CCGCACTGGAAGAGTCTGAGCA	AGCCGCTTCTT	8572
QY	1952	GCTGGGCACTGTGCTATGAG	CAGAGGCCCAAGCAGCTGATG	AGAGGCGGAGCGGATTT	2011
Db	8574	GCTGGGCACTGTGCTATGAG	CAGAGGCCCAAGCAGCTGATG	AGAGGCGGAGCGGATTT	8633
QY	2012	CTACGAGCAGAAACAGCCGTAC	CACTACCTTTGTGGGGGCGAGAA	GGCAACACTCTGGA	2071
Db	8634	CTACGAGCAGAAACAGCCGTAC	CACTACCTTTGTGGGGGCGAGAA	GGCAACACTCTGGA	8693
QY	2072	CTATGCCAACAAGCAGCTGAG	CGGGGTGTGTGTGCGCAACTA	CTACACCCCTGTGTGCGGCT	2133
Db	8694	CTATGCCAACAAGCAGCTGAG	CGGGGTGTGTGTGCGCAACTA	CTACACCCCTGTGTGCGGCT	8753
QY	2132	TTTCTGTGAGGGCGCTGTGTGA	CAGGTGTGGCCCAAGGGCATC	CCCTTCCAAACAGACCAAGTT	2197
Db	8754	TTTCTGTGAGGGCGCTGTGTGA	CAGGTGTGGCCCAAGGGCATC	CCCTTCCAAACAGACCAAGTT	8813
QY	2192	TGACAAATAATGTCTTCCAA	CTGAGACAGAGCCTTCGTTCT	CTAGCAAGAGAGTACCCGAG	2251
Db	8814	TGACAAATAATGTCTTCCAA	CTGAGACAGAGCCTTCGTTCT	CTAGCAAGAGAGTACCCGAG	8872
QY	2252	CCAGCCGCGAGAGACACTGTG	TGGACTGTGGCCCAAGAAATCTT	CCTCAAAATATTAACCCGCG	2311
Db	8874	CCAGCCGCGAGAGACACTGTG	TGGACTGTGGCCCAAGAAATCTT	CCTCAAAATATTAACCCGCG	8933
QY	2312	CTGGGTGTGCGCGCTCTTGTG	TGATTAATTCGACACACTGTGG	CCCTTGTTTCCGCTAATTC	2371
Db	8934	CTGGGTGTGCGCGCTCTTGTG	TGATTAATTCGACACACTGTGG	CCCTTGTTTCCGCTAATTC	8993
QY	2372	CAGGGCAGATTCAGGGGCCCA	AGCTGTGACAGACATCA	CAGAGATTAACCAAGGCTGTGGAG	2433
Db	8994	CAGGGCAGATTCAGGGGCCCA	AGCTGTGACAGACATCA	CAGAGATTAACCAAGGCTGTGGAG	9053
QY	2432	GAGAGCCCAAGGCGCTGTGTG	TGGAGGTCTGTGACTGTGGGGGAA	TGTGAGGGAAATATACCTGCG	2497
Db	9054	GAGAGCCCAAGGCGCTGTGTG	TGGAGGTCTGTGACTGTGGGGGAA	TGTGAGGGAAATATACCTGCG	9113
QY	2492	CTCCACACCAACCCAAAGTGTG	GGGATTAAGTACTGTTTCTTTCC	ACTTTAAA	2544
Db	9114	CTCCACACCAACCCAAAGTGTG	GGGATTAAGTACTGTTTCTTTCC	ACTTTAAA	9166

[illegible]

```

QY 1783 TGAACCTCACTGCGAGGAGCTGACAGAGCTGCTGATGATAGAGGAGCAAGA 1842
DB 661 TGAACCTCACTGCGAGGAGCTGACAGAGCTGCTGATGATAGAGGAGCAAGA 720
QY 1843 GGGCTTACTGAGCAAGAGAGCTGCTGCTGCTGCTGAGAGGCTGCTGCTGAG 1902
DB 721 GGGCTTACTGAGCAAGAGAGCTGCTGCTGCTGCTGAGAGGCTGCTGCTGAG 780
QY 1903 AGCTGCTGCGGCACTGAGCAGAGTGTGCTGAGTGAAGAGCGCTTCTTGCTGGGCAAGCT 1962
DB 781 AGCTGCTGCGGCACTGAGCAGAGTGTGCTGAGTGAAGAGCGCTTCTTGCTGGGCAAGCT 840
QY 1963 GGTAGAGCAGGCGCGAGAGCAGGAGCTGAGTGAAGAGCGCGAGCTTCTGAGAGCA 2022
DB 841 GGTAGAGCAGGCGCGAGAGCAGGAGCTGAGTGAAGAGCGCGAGCTTCTGAGAGCA 900
QY 2023 ACAGCGGCTACAGCAGCTTGTGAGGAGCGAGAGCAATCTGAGCTATGCGCAACA 2082
DB 901 ACAGCGGCTACAGCAGCTTGTGAGGAGCGAGAGCAATCTGAGCTATGCGCAACA 960
QY 2083 AGCAGCTGCGGAGGTGTGAGCGCACTACTACACCCCTGCTGAGCGCTTTTCTGAGAG 2142
DB 961 AGCAGCTGCGGAGGTGTGAGCGCACTACTACTACACCCCTGCTGAGCGCTTTTCTGAGAG 1020
QY 2143 CGCTGCTTGAAGTGTGAGCGCGAGCGCTTCTTCTGAGCAGAGCTTGAAGCAAAATG 2202
DB 1021 CGCTGCTTGAAGTGTGAGCGCGAGCGCTTCTTCTGAGCAGAGCTTGAAGCAAAATG 1080
QY 2203 TCTTCAACTGAGAGAGCGCTTCTTCTGAGCAGAGGTAACCCGAGCGCGAG 2262
DB 1081 TCTTCAACTGAGAGAGCGCTTCTTCTGAGCAGAGGTAACCCGAGCGCGAG 1140
QY 2263 GAGACACTGTGAGCTGAGCGCAAGATCTTCTCAAAATTTACCCGCGTGGGTGCGG 2322
DB 1141 GAGACACTGTGAGCTGAGCGCAAGATCTTCTCAAAATTTACCCGCGTGGGTGCGG 1200
QY 2323 GCTCTTGTGTA 2333
DB 1201 GCTCTTGTGTA 1211

```

```

RESULT 4
US-10-027-632-111960
; Sequence 111960, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111960
; LENGTH: 2796
; TYPE: DNA
; ORGANISM: Human

```

```

US-10-027-632-111960
Query Match 18.9%; Score 485.6; DB 13; Length 2796;
Best Local Similarity 85.9%; Pred. No. 4,8e-114;
Matches 550; Conservative 0; Mismatches 89; Indels 1; Gaps 1;
QY 1 CCGGAGCTTACGCTTGGGTGCAAGTGTGAGCGAGAGCGAGCTGATGAGCGAGCGG 60
DB 1017 CCGGAGCTTACGCTTGGGTGCAAGTGTGAGCGAGAGCGAGCTGATGAGCGAGCGG 1075
QY 61 CCGGAGCTTACGCTTGGGTGCAAGTGTGAGCGAGAGCGAGCTGATGAGCGAGCGG 120
DB 1076 CCGGAGCTTACGCTTGGGTGCAAGTGTGAGCGAGAGCGAGCTGATGAGCGAGCGG 1135
QY 121 CCGGAGCTTACGCTTGGGTGCAAGTGTGAGCGAGAGCGAGCTGATGAGCGAGCGG 180
DB 1136 CCGGAGCTTACGCTTGGGTGCAAGTGTGAGCGAGAGCGAGCTGATGAGCGAGCGG 1195
QY 181 GGGAGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
DB 1196 GGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1255
QY 241 ACTTCTCCGCTGCTGAGAGCGGCTCTGAGCTGCAAGCGAGCGGCTTGAACCTACAGCC 300
DB 1256 ACTTCTCCGCTGCTGAGAGCGGCTCTGAGCTGCAAGCGAGCGGCTTGAACCTACAGCC 1315
QY 301 TGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 360
DB 1316 TGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1375
QY 361 CGGAGCTGACCGCTACCTGCGGAGCTTGTGAGCTGCAAGCGGCTTGAACCTACAGCC 420
DB 1376 CGGAGCTGACCGCTACCTGCGGAGCTTGTGAGCTGCAAGCGGCTTGAACCTACAGCC 1435
QY 421 AGCTGCGCTGCGCGCGGCACTGCGAGCGGCTGCGGAGAGCTGACCGAGCGCGCA 480
DB 1436 AGCTGCGCTGCGCGCGGCACTGCGAGCGGCTGCGGAGAGCTGACCGAGCGCGCA 1495
QY 481 ACAGGTACCGCTTACAGAGATGTGACAGCGCAAGCTACCTTCTGCTGTGGAGCT 540
DB 1496 ACAGGTACCGCTTACAGAGATGTGACAGCGCAAGCTACCTTCTGCTGTGGAGCT 1555
QY 541 GGGCGGCTGAGAGCGAGAGATAGTGTGAGCGGCTGATGAGCGAGCTGAGCGG 600
DB 1556 GGGCGGCTGAGAGCGAGAGATAGTGTGAGCGGCTGATGAGCGAGCTGAGCGG 1615
QY 601 CCGGAGCGGCGGAGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGG 640
DB 1616 CCGGAGCGGCGGAGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGG 1655

```

```

RESULT 5
US-10-027-632-111960
; Sequence 111960, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28

```

```
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11960
; LENGTH: 2796
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-111960

Query Match      18.3%; Score 485.6; DB 14; Length 2796;
Best Local Similarity 85.3%; Pred. No. 4.8e-114;
Matches 550; Conservative 0; Mismatches 89; Indels 1; Gaps 1;

QY 1 CCGGGGCTTACCTGCGGCTGACGATGAGCGGCGGAGCTGATGAGCGGCGGCGG 60
DB 1017 CCGGGGCTTACCTGCGGCTGACGATGAGCGGCGGAG-GCCGCGAGCTGATGAGCGGCGGCGG 1075
QY 61 CCCCACCCCTGCGCGCTGCGGAGACCCGAGACTGAGACATGAGAGCGGCTGCGGCTG 120
DB 1076 CCCCACCCCTGCGCGCTGCGGAGACCCGAGACTGAGACATGAGAGCGGCTGCGGCTG 1135
QY 121 CCGCGGCGGCTGCGGCTGCTTCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 1136 CCGCGGCGGCTGCGGCTGCTTCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1195
QY 181 GGGAGGCGGCGGCGGCTGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
DB 1196 GGGAGGCGGCGGCGGCTGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1255
QY 241 ACTTCTCCGCTGCGGCTGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
DB 1256 ACTTCTCCGCTGCGGCTGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1315
QY 301 TGGGCGGCGGCGGCGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
DB 1316 TGGGCGGCGGCGGCGGCGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1375
QY 361 CCGGCGGCTGACCGCTACCTGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCTG 420
DB 1376 CCGGCGGCTGACCGCTACCTGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCTG 1435
QY 421 AGCTGCGGCTGCGGCGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
DB 1436 AGCTGCGGCTGCGGCGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1495
QY 481 ACAAGTACCGCTATTAACAGAAATGTGTGACCGCAAAAGCTACTCTTGTGTGTGAGCT 540
DB 1496 ACAAGTACCGCTATTAACAGAAATGTGTGACCGCAAAAGCTACTCTTGTGTGTGAGCT 1555
QY 541 GGGCGCGGCTGCGGCGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
DB 1556 GCGGCGGCGGCGGCGGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1615
QY 601 CCGGAGCGGCGGCGGAGGCGGCTGCGGCGGCGGCTGCTT 640
DB 1616 CCGGAGCGGCGGCGGAGGCGGCTGCGGCGGCGGCTGCTT 1655

RESULT 6
US-10-027-632-141597
; Sequence 141597, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; PRIOR APPLICATION NUMBER: US/10/027, 632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218, 006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198, 676
```

```
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 141597
; LENGTH: 808
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-141597

Query Match      17.7%; Score 456.6; DB 13; Length 808;
Best Local Similarity 99.8%; Pred. No. 1e-106;
Matches 456; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2088 CTGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2147
DB 1 CTGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
QY 2148 GTTACAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2207
DB 61 GTTACAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
QY 2208 CACTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2267
DB 121 CACTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
QY 2268 ACTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2327
DB 181 ACTGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
QY 2328 TGTGATGATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2387
DB 241 TGTGATGATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
QY 2388 GCCAGAGCTGAGACAGATCATGATTAACCCAGGCGGCGGCGGCGGCGGCGGCGGCGG 2447
DB 301 GCCAGAGCTGAGACAGATCATGATTAACCCAGGCGGCGGCGGCGGCGGCGGCGGCGG 360
QY 2448 CTGGTGGGCTGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2507
DB 361 CTGGTGGGCTGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
QY 2508 AGTGGGATTAAGTACTGTTTCTTCCACTTAA 2544
DB 421 AGTGGGATTAAGTACTGTTTCTTCCACTTAA 457

RESULT 7
US-10-027-632-141597
; Sequence 141597, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; PRIOR APPLICATION NUMBER: US/10/027, 632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218, 006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198, 676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193, 483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185, 218
```

```

; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 141597
; LENGTH: 808
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-141597

Query Match      17.7% Score 456.6; DB 14; Length 808;
Best Local Similarity 99.8%; Pred. No. 1e-106;
Matches 456; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2088 CTGGCGGGGTTGGTGGCCCACTACTACACCCCTGCTGGCGGCTTTTCTTGGAGGCGCTG 2147
DB 1 CTGGCGGGGTTGGTGGCCCACTACTACACCCCTGCTGGCGGCTTTTCTTGGAGGCGCTG 60
QY 2148 GTTACAGTGTGGCCGAGGAGATCCCTTCCAGAGACAGATTTGACAAAATGCTTC 2207
DB 61 GTTACAGTGTGGCCGAGGAGATCCCTTCCAGAGACAGATTTGACAAAATGCTTC 120
QY 2208 CAACGTGAGCAGGCGCTTGTCTCAGACAGAGAGTACCAGCCAGCCGCGAGAGAC 2267
DB 121 CAACGTGAGCAGGCGCTTGTCTCAGACAGAGAGTACCAGCCAGCCGCGAGAGAC 180
QY 2268 ACTGTGACCTGGCCAGAAATCTTCTCAATATTAACCCGGCTGGTGGCGGCTCT 2327
DB 181 ACTGTGACCTGGCCAGAAATCTTCTCAATATTAACCCGGCTGGTGGCGGCTCT 240
QY 2328 TGTGATGATTCGCCACCACTGGGCGCTTTTCCCGCTAATTCAGAGGAGATTCAGG 2387
DB 241 TGTGATGATTCGCCACCACTGGGCGCTTTTCCCGCTAATTCAGAGGAGATTCAGG 300
QY 2388 GCCCAGAGCTGGACAGACATCACAGATTAACCCGCTGGAGAGAGGCCCCACGCGCTG 2447
DB 301 GCCCAGAGCTGGACAGACATCACAGATTAACCCGCTGGAGAGAGGCCCCACGCGCTG 360
QY 2448 CTGATGGGGTGTGACCTGGGGGAGTTGGAGGGAATACCTGCCCTGCACACCAACCAA 2507
DB 361 CTGATGGGGTGTGACCTGGGGGAGTTGGAGGGAATACCTGCCCTGCACACCAACCAA 420
QY 2508 AGTGTGGATTAAGTACTGTTTCTTTCCACTTAA 2544
DB 421 AGTGTGGATTAAGTACTGTTTCTTTCCACTTAA 457

RESULT 8
US-09-918-995-30489
; Sequence 30489, Application US/09918995
; Publication No. US2003007623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 30489
; LENGTH: 499
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: (1)...(499)
; OTHER INFORMATION: n = A,T,C or G
; US-09-918-995-30489

Query Match      17.3% Score 445.4; DB 11; Length 499;
Best Local Similarity 99.8%; Pred. No. 6.7e-104;
Matches 446; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1169 CCAGCCGACAGTCTGCGGGGCGCCGACAGATCAAGGCTGTGAGAGCTGTGCCCCCGGG 1228
DB 53 CCAGCCGACAGTCTGCGGGGCGCCGACAGATCAAGGCTGTGAGAGCTGTGCCCCCGGG 112
QY 1229 CCGCTCTGTTCTGAGACTGTGTTGCTGAGAGCCAGCTGTATACCCGACCTGCTC 1288
DB 113 CCGCTCTGTTCTGAGACTGTGTTGCTGAGAGCCAGCTGTGTATACCCGACCTGCTC 172
QY 1289 CTTCCAGGCGCAGCCCTTCACTGTGTGATGTGTGACAACTTTGGGGGAAACATGCTCT 1348
DB 173 CTTCCAGGCGCAGCCCTTCACTGTGTGATGTGTGACAACTTTGGGGGAAACATGCTCT 232
QY 1349 TTTTGAGCCCTTGAAGAGCTGTGACAGAGGCCAGAGCTGCCCTTTCCCACTC 1408
DB 233 TTTTGAGCCCTTGAAGAGCTGTGACAGAGGCCAGAGCTGCCCTTTCCCACTC 292
QY 1409 CACCATGTGAGCAGAGGCTGTGACAGAGGCCATCAGCCAGAACGAGTGTCTATTTC 1468
DB 293 CACCATGTGAGCAGAGGCTGTGACAGAGGCCATCAGCCAGAACGAGTGTCTATTTC 352
QY 1469 CCTCATGCTGAGCTGGGCTGGCGAAAGAACCCAGTGCAGATTGGCAGCTGGGTGAC 1528
DB 353 CCTCATGCTGAGCTGGGCTGGCGAAAGAACCCAGTGCAGATTGGCAGCTGGGTGAC 412
QY 1529 CAGCTTTGCCCGCGCGGTATGAGGCTCTCCACCTCGACAGAGGGGACGCTGAGAGCT 1588
DB 413 CAGCTTTGCCCGCGCGGTATGAGGCTCTCCACCTCGACAGAGGGGACGCTGAGAGCT 472
QY 1589 ACTGCTCGGAGTGTGTCAACTGCTC 1615
DB 473 ACTGCTCGGAGTGTGTCAACTGCTC 499

RESULT 9
US-10-029-386-9071
; Sequence 9071, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEWICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9071
; LENGTH: 522
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO U34879.1
; OTHER INFORMATION: EXPRESSED IN HBLA, SIGNAL = 0.74
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: NT HIT: U34879.1, EVALUATE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P54802, EVALUATE 2.80e-36
; OTHER INFORMATION: EST_HUMAN HIT: AM205518.1, EVALUATE 0.00e+00
; US-10-029-386-9071

Query Match      16.5% Score 426; DB 13; Length 522;
Best Local Similarity 100.0%; Pred. No. 6.1e-99;

```

Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2119 CTGGCTGGGGCTTTTCTTCTGGAGGCGCGCTGTGAAGAGTGGCCCAAGGCGATCCCTTTCC 2178  
DB 1 CTGGCTGGGGCTTTTCTTCTGGAGGCGCGCTGTGAAGAGTGGCCCAAGGCGATCCCTTTCC 60

QY 2179 AACAGACCACTTTGACAAAATGTCTTCCACTGAGAGAGGCTTCTTCTCAGCAAGC 2238  
DB 61 AACAGACCACTTTGACAAAATGTCTTCCACTGAGAGAGGCTTCTTCTCAGCAAGC 120

QY 2235 AAGAGTACCCCGACCGCCGCGAGAGAGACTGTGACCTGTGGCCAAAGAGATCTTCCCA 2298  
DB 121 AAGAGTACCCCGACCGCCGCGAGAGAGACTGTGACCTGTGGCCAAAGAGATCTTCCCA 180

QY 2299 AATATTACCCCGCTGGGTGGCGCGCTGTGTGATAGATTGCCACCACTGGGCTTGT 2358  
DB 181 AATATTACCCCGCTGGGTGGCGCGCTGTGTGATAGATTGCCACCACTGGGCTTGT 240

QY 2359 TTTCCGCTAATTCAGAGGAGATTCAGAGGCGCCAGAGCTGACACATCAAGATTAAC 2418  
DB 241 TTTCCGCTAATTCAGAGGAGATTCAGAGGCGCCAGAGCTGACACATCAAGATTAAC 300

QY 2419 CCAGGCTTGGAGAGAGGCGCCCAAGGCTGTGGTGGGCTGTGACCTGGGGGATTTGAGG 2478  
DB 301 CCAGGCTTGGAGAGAGGCGCCCAAGGCTGTGGTGGGCTGTGACCTGGGGGATTTGAGG 360

QY 2479 GAAATGACCTGCTTCCACACACCAACCAAGTGTGGGATTTAAAGTACTGTTTCTTCCA 2538  
DB 361 GAAATGACCTGCTTCCACACACCAACCAAGTGTGGGATTTAAAGTACTGTTTCTTCCA 420

QY 2539 CTTAAA 2544  
DB 421 CTTAAA 426

RESULT 10  
US-09-918-995-24382  
; Sequence 24382; Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 24382  
; LENGTH: 449  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-24382

Query Match 15.3%; Score 395; DB 11; Length 449;  
Best Local Similarity 99.8%; Pred. No. 5e-91;  
Matches 406; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2150 TGACAGTGTGGCCCAAGGCGCTTCTTCCAAAGCAACCACTTGAAGAAAATGTCTTCCA 2209  
DB 33 TGACAGTGTGGCCCAAGGCGCTTCTTCCAAAGCAACCACTTGAAGAAAATGTCTTCCA 91

QY 2210 ACTGAGACAGGCTTCTGCTTCTCAGACAGAGTACCCCAAGCCGAGCGAGAGACAC 2269  
DB 92 ACTGAGACAGGCTTCTGCTTCTCAGACAGAGTACCCCAAGCCGAGCGAGAGACAC 151

QY 2270 TGTGACCTTGGCCAAAGATCTTCTCAATATTAACCCGGGTGGGTGGCGGCTCTTG 2329  
DB 152 TGTGACCTTGGCCAAAGATCTTCTCAATATTAACCCGGGTGGGTGGCGGCTCTTG 211

QY 2330 GTGATAGATTGCGACCACTGGGCTTGTCTTCCGCTAATTCAGGAGCATTTCCAGGCG 2389

DB 212 GTGATAGATTGCGACCACTGGGCTTGTCTTCCGCTAATTCAGGAGCATTTCCAGGCG 271

QY 2390 CCAGAGCTGGAGACAGATCAACAGATTAACCAAGGCTTGGAGAGAGGCGCCGCTGCT 2449  
DB 272 CCAGAGCTGGAGACAGATCAACAGATTAACCAAGGCTTGGAGAGAGGCGCCGCTGCT 331

QY 2450 GGTGGGCTTGAACCTTGGGGGATTTGAGGGAATGACTGCTTCCACCAACCAAG 2509  
DB 332 GGTGGGCTTGAACCTTGGGGGATTTGAGGGAATGACTGCTTCCACCAACCAAG 391

QY 2510 TGTGGATTTAAGTACTGTTTCTTCCACTTAAAAA 2556  
DB 392 TGTGGATTTAAGTACTGTTTCTTCCACTTAAAAA 438

RESULT 11  
US-09-918-995-21830  
; Sequence 21830; Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 21830  
; LENGTH: 468  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(468)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-21830

Query Match 13.9%; Score 358.4; DB 11; Length 468;  
Best Local Similarity 99.2%; Pred. No. 1.e-81;  
Matches 359; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2195 CAAAATGTCTTCCAACTGAGAGAGGCTTCTTCTCAGACAGAGTACCCAGCCA 2254  
DB 29 CAAAATGTCTTCCAACTGAGAGAGGCTTCTTCTCAGACAGAGTACCCAGCCA 88

QY 2255 GCCCGAGAGACACTGTGACCTGGCCAAAGATTTCTCAATATTACCCCGCTG 2314  
DB 89 GCCCGAGAGACACTGTGACCTGGCCAAAGATTTCTCAATATTACCCCGCTG 148

QY 2315 GGTGGCGGCTCTTGTGATAGATTGCCACCACTGGGCTTGTCTTCCGCTAATTCAG 2374  
DB 149 GGTGGCGGCTCTTGTGATAGATTGCCACCACTGGGCTTGTCTTCCGCTAATTCAG 208

QY 2375 GGCAGATTCCAGGGCCCAAGCTTGAACATCAAGATTAACCAAGGCTTGGAGAG 2434  
DB 209 GGCAGATTCCAGGGCCCAAGCTTGAACATCAAGATTAACCAAGGCTTGGAGAG 268

QY 2435 GCCCAAGGCTGTGTGTGGGGTCTGACCTGGGGGATTTGAGGGAATGACCTGCCCTC 2494  
DB 269 GCCCAAGGCTGTGTGTGGGGTCTGACCTGGGGGATTTGAGGGAATGACCTGCCCTC 328

QY 2495 CACCAACCAACCAAGTGTGATTAAGTACTGTTTCTTCCACTTAAAAA 2554  
DB 329 CACCAACCAACCAAGTGTGATTAAGTACTGTTTCTTCCACTTAAAAA 388

QY 2555 AA 2556  
DB 389 AA 390



QY	1793	TCGGCAGGAGTACGAGAGCTGTCAGTTGTACTAGAGAGGCAAGAGCGCTTACT	1855
Db	1665	CCGCCAGGACATCTTCCAAACCGACGCGGGATACCTCTCCCGCAATACGAGACGCGTACGA	1744
QY	1853	GAGCAAGAGAGCTGCGCCTCCCTGTTTGAAGGAGCGAGCGTCTGGCCTATGAGCTGTGC	1912
Db	1725	GCGCAAGGACACCGCCCGCTTGACCGGCTGACCGGCGTGTGG---CTGGCCCTATGGA	1781
QY	1913	GGCATGACGAGGTGCTGGCTAGTACAGCGCGTTCTTGCTGGGCACTGCTAGACA	1972
Db	1782	CTGTGAGAGGACATCTCTGGCCACCGACTCCCGTCACTGCTGGGCGGTTGGGTGGCGA	1841
QY	1973	GGCGCCGAGCAGCGGCGAGTCACTGAGAGGCGGAGCGGATTTTAAAGACAAACAAGCGCTA	2032
Db	1842	CGCGCGGCGTGGGAGTGCAGCGCGCGGAGACCGGACCGGCTGCGTACGACGCGCTTCTC	1901
QY	2033	CCAGCTGACCTTGTG-----GGGGCCAGAGAGCAACATCCTGAGACTATGCCAA	2080
Db	1902	GCTGCTGACCGCTGTGGGGGACACCGCGGGGTGCGGAGCGCGGGCTCGGACCTAGCGCCA	1961
QY	2081	CAAGCAGCTGGCGGGGTTGTGTGGCCACTACTACACCCCTCGCTGGCGGCTTTCTGGA	2140
Db	1962	CCGCGAGTGGCGGGACTGATCGGCGGGCTGTACCGGCTCGCTGTGTCAGACGACTTCGC	2021
QY	2141	GGCGCTG 2147	
Db	2022	GGAGCTG 2028	

RESULT 13  
US-10-156-761-1

Sequence 1, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMTA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIDA  
TITLE OR INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 1  
LENGTH: 9025608  
TYPE: DNA  
ORGANISM: Streptomyces avermectilis  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (4157715)  
OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

```

Query Match Similarity 7.1% Score 13.4; DB 15; Length 9025608;
Best Local Similarity 46.9% Pred. No. 3,8e-36;
Matches 961; Conservative 0; Mismatches 996; Indels 90; Gaps 9;

QY 140 TCTCTGAGCGGGAGCGCGGAGCGAGGACAGAGCGCGGAGCGCGGCGTGC 199
Db 2453665 TCTCTCCGCGGGATCGCGGAGCGCGCGCTGCTGTACC--CGGCTGAGCGGACG 2453721

QY 200 GCGCGCTGTGAGCCGAGCTGTGAGGACAGAGCCCGAGGACGACTTCTCGTGTGCGTGA 259
Db 2453722 AGCGCGCGGAGGAGGCTCGGCGCGGACGAGCGCGCGCGGAGCTCTCTCCGCTCACTGCG 2453781

```

QY	260	GC	CGCGCTCTGAGCTGCCAAGCCGGGCTTTGGACAACTTCAACCTCTGGGGCGGGCGCGCGGCG	319
Db	2453782	GC	AGCTCAACGTTCCGCAACGGCCGGGCGGACCGGACAGTTTCAAGTGTCCGGCCGTAACGGG	2453841
QY	320	GC	CGGCTCGGGGTGCGCGGCTCCACCGGAGCGGCGGCGCGCGCGCGCGCGGCTGCACCGCTACCT	379
Db	2453842	CC	GGTCAACCGTCAACGGGCGGGACACACCGGGACCCAGGCTCAGGGGGCTGAACCTGTAATT	2453901
QY	380	GC	CGCAGCTTCTGTGGCTGCCAGGTGGCTGTGGTCCGGCTCTCACTGGCGCTGTCCGGCGCC	439
Db	2453902	GAG	GAATGATGCGGACCCCACTCAACTGGGGGGGCAAGCACTCCGCTGCCGCGC	2453961
QY	440	ACT	CCGACCGCGTGGCGGGGAGCTGACCGAGGCGCACGCCCAAGGTTACCGCTATTACCA	499
Db	2453962	ACT	CGCGGCGCTGCACCGGACAGTGAACCGGCGCGGAAAGTCCCGACCGCTTCGCGCT	2454021
QY	500	GA	ATGTGTGACCGCAAGTACTCTCTTGTGTGGTGGAGTGGGCCCGCTGGAGCGAGA	559
Db	2454022	CA	AGACACGACGACCGGCTACCGGGCCGTACCAAGTCAAGTCACTGGAGAGGTGA	2454081
QY	560	GAT	GACTGATGGCGCTGGAATGGACATCACTGGACCTGGCGCGGCGGACGAGCG	619
Db	2454082	GCT	GACGTAATGGCCCTGACCGGCTACCAAGAGTGTGTCTCAACGGGCGCGAGCG	2454141
QY	620	CAT	CTGCACGCGGTGTACTGGCTTGGGCTTGGGCTGACCCAGGACAGATCAATGATTCTT	679
Db	2454142	GCT	CAACACCGGGTGTCCAGAGGATTCCGGTACACCGACGAGAGGTGCGAAGTGGAT	2454201
QY	680	TAC	TGTGTCTGTGCTCTCTTGCGCTTGGGGGCGAATGGGCACTTGACACCTGGGATGGCGCC	739
Db	2454202	TCC	GGGCGCGGACCATGAGCCCTGTGGTGGCTGTGCAGAACTCTTCGCTTCCCGACCC	2454261
QY	740	CCT	CGCCCTCTCTGGACATCAAGCAGCTTTACTGTGCACACCGGATCCTTGACAGAT	799
Db	2454262	CGT	CTCAACAAGTCTCTGACCGCCGGGCGCGGCTTGCGCGGGGATTCGCAACCGGCT	2454321
QY	800	GC	GTCTCTTGGGATGACCCCAAGTGTCTGCAATTCGCGGGGATGTTCCGAGGCTGT	859
Db	2454322	GC	GGAGCTGGGACATGACACCGGTTTCCCGGCTACTTTCGGACAGGTTTCGCGCGCTT	2454381
QY	860	CAC	CAAGGTTTCCCTAGGTCATGTACGAAAGATGGGAGTGGGGGCACTTTAACTG	919
Db	2454382	CG	CGGACCGGACCGGGGTGCGCACACCTGTCCGAGGGACCTGATGGGCTTCCGACG	2454441
QY	920	TT	CTTACTCTGTCTCTCTCTCTTCTGTGGGTCCGAAAGCCCAATTCCTCCATCATCGGAG	979
Db	2454442	GCC	GACGATCGG-----CTGGACCGGCGCACCGACACTTCAACGGGTGGCGCGGC	2454492
QY	980	CCT	CTTCTCTGGGAAGCTGATCAAAAGTTTGGACAGACCAATCATGATGGGGCGGACAC	1039
Db	2454493	CTT	CTACCGGATCCAGGACGAGAGTGTGGGGGGCTTCGACCCGCTTCAAAATGAGCT	2454552
QY	1040	TTT	CAATGATGAGGCACCTTCTCTAGAGCCCTCTTACTTTCGCGAGCCACCACTGC	1099
Db	2454553	GCT	GACGAGGGCGGACCGCCCGGATCGTGGTGTACTCTGGGCTGAG	2454606
QY	1100	CGT	CTATGAGGCATGACTGCAGTGTACTGAGGCTGTGTGGTGTCTCCAGGCTGGCT	1159
Db	2454607	CGT	CGAGAGGCGCTTGGCGGGCGGACACCGGGGCGGTAATGAGTACTCTGGGCTGGC	2454665
QY	1160	CTT	CCAGACAGCGGAGTGTGGGGGCGCGCCCAATCAAGGCTGTGTGGAGGTGT	1219
Db	2454666	-----	AGCAACAACCGGCCAGGGCGATTCGCGACGCGGT	2454699
QY	1220	GCC	CGTGGCGCTCTGGTCTTGAACCTGTGTCTGAGAGCGACGCTGTGTATAC---	1276
Db	2454700	GG	ACAGGACCGGATGCTGTGTGTGAGACGCGCTCTGTGACCGCTTCCGAAAGTCAACGA	2454759
QY	1277	CCG	CACTGCCCTCTTTCAGGGGCGACCCCTTCACTGGGATGTGTGACAACTTTGGGGG	1336
Db	2454760	CCG	CGAGGCGGACCTGGCACCGGACCCCGTATGCTTGGCTGTGATCTTGAACTTCCGCGG	2454819
QY	1337	AA	ACCATGATCTTTTGGAGCCTTAGAGGCTGTGAAACGAGGCCCAAGACCTGCCGCT	1396

```

Db 2454820 CCAATGACCTGCGGCGGAAACACTCCGACTGAGCTTGATGACGAGCGGTGCGCAC 2454879
Qy 1397 CTTCCCACTCCACCATGATAGGACAGGACATGAGCCCGGAGGACATCAGCAGAACGA 1456
Db 2454880 CAGGCGGCGGACACCTCCGCGGTGTGTCTGTCTCCGAGGCGCGGACAAACATCC 2454939
Qy 1457 AGTGTCTATTCCCTCATAGCTGAGCTGGGCTGGCGAAAGAACCAAGTGCAGATTGGC 1516
Db 2454940 GAGCGCTTCGCGCTGTTTTCGAACTCGCTGGCGCAGGCGCACTTG---GACTGCG 2454996
Qy 1517 AGCCGTGGTGAACAGCTTTCGCGCGCGGTATGGGGTCTCCACCGGACGAGCGAGGCG 1576
Db 2454997 CGCATGTTTCGCGCGGTGGCGCGCTCGCGCTACGCGGAGACCATTCACATGCGAGGC 2455056
Qy 1577 AGCGTGAAGCTACTGCTCCGAGAGTGTATACATGCTCCGAGGAGGCGCTTGACAGGCGCA 1636
Db 2455057 CGCCTGGACATCTCGCGCGGACCGGTACGCGACACGCGCGCGCATCTGTGAGCGA 2455116
Qy 1637 CAATGTAGCCCGCTGTGACAGCGCGCTCCCTACAGATGAATACA----- 1683
Db 2455117 GGGCGGAGCGGCTGTTCGGGCGCGCGCTCCCTCGCGCGACGAAAGCGCTTCTG 2455176
Qy 1684 -----GCATCTGTATCAACCGATCTGATGTGTTGAGGCTTGCGGCTGTGCT 1732
Db 2455177 GTCCGCGAAGCGCTCCGTTACCGGCGCGAGAGTTCGAAACCGCGCTTGCGGAGCTCT 2455236
Qy 1733 CACATCTGCTCTCTCTCTGCGCACACCGCCGCTTCCGCTTACGACTGTGAGCTTAC 1792
Db 2455237 GAAAGTGGCGCGCGGCGGTGCGCGGCTCGTCCGCGTACCGCGCACCTGCTGACGTGCG 2455296
Qy 1793 TCGGACGAGAGTGAAGAGCTGTGTGACTGTATGAGGAGGAGCAAGAACCGCTTACT 1852
Db 2455297 CGCGAGGACATCTTCAACCGACCGGCTACTCTCCCGAGATCAGAGCGGCTTACGA 2455356
Qy 1853 GAGCAAGAGCTGGCTCTCTCTGTTGAGGCGTGAAGGCGTCTGCTTATGAGCTGCGC 1912
Db 2455357 GAGCAAGAGCACCGCGCTTCAACCGGCTGACCGGCGCTGCG---CTGGCGCTGATGGA 2455413
Qy 1913 GGCACGTGACAGAGTGTGCGGTATGAGACGCGCTTCTTGCGGACGCTGAGTGAAGCA 1972
Db 2455414 CCTGTGAGAGCACTCTCGGCGCACCGACTCCGCTCATCTGTGCGCGGTGTGGTGGCGCA 2455473
Qy 1973 GCGCCGAGCAGCGGAGTCAAGTGAAGCGGACGATTTTTCAGAGCAAGAACCGCTA 2032
Db 2455474 CGCGCGGCGGTGGGTGCGAGGCGGCGGAGCGGACCGGCTCGGTACGACGCGCTTC 2455533
Qy 2033 CCACTTACCTTTG-----GGGCGAAGAGGCAACATCTGACTATGCCAA 2080
Db 2455534 GCTGTGACCGTGTGGGACACCGCGCGGCGGACCGCGGCTGCGCACTACGCGCAA 2455593
Qy 2081 CAAGCAGCTGGCGGGTGTGTGCACTACTACACCGCTTGCGGCTTTTCTGGA 2140
Db 2455594 CGCGAGTGGGCGGACTGTGTGCGGCGGTATCCGCGTGTGAGTACTTTCG 2455653
Qy 2141 GCGCGTG 2147
Db 2455654 GAGCTG 2455660

```

```

RESULT 14
US-10-156-761-5965
; Sequence 5965, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, HIROSHI
; APPLICANT: HORIKAWA, JUN
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

```

```

; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5965
; LENGTH: 3114
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3114)
; US-10-156-761-5965

```

```

Query Match 6.0%; Score 155; DB 15; Length 3114;
Best Local Similarity 51.7%; Pred. No. 1,6e-29;
Matches 353; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

```

```

Qy 178 CCGGAGCGCGCGCGCTGCGGCGCTGTGACCCCGCTGTGGGCGAGCCCGCGG 237
Db 101 CCGGAGGAGCGCGCGCTTGTATACGCTCCGACGCTCCGCTCATCGGCTGTGCG 160
Qy 238 CGGACTTCCGTTGCGTGAAGCGCGCTGTGCGGCGCAAGCGGCGTTGACACTTACA 297
Db 161 CCGGAGATGCGGACATTCGCTTACGCTTGTGTGCGGACCGGCGCGGCGGACCTT 220
Qy 298 GCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 357
Db 221 TCGGCGTCAACCGCGCGCACAGGCGCGCATCAGGTGTGCGGTACGACCGGCGGTGCG 280
Qy 358 CCGCGGCGTGAACCGCTACTGCGGAACTTGTGTGCGTGTGCGGCGGCGGCGGCGG 417
Db 281 TCAACCGGCGTCACTGTGACTGTGAAGTACGTGTGCGGCGGCGGCGGCGGCGGCGG 340
Qy 418 CTAAGTGGCGCTGCGCGCGCGCACTGCGACCGGTGCGGCGGAGGCTGACGAGCGCACG 477
Db 341 GCGAGTGAACCTGCGCGCGGAGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 400
Qy 478 CCAACAGTACCGCTATTACAGAAATGTGTGACGCGAAGCTTACTCTTGTGTGTGG 537
Db 401 CCTGTCCACCGCTTTCGCGCTCAACGACCAACGACGCGCTTACAGGCGCGCTACGCG 460
Qy 538 ACTGGCGCGCGTGGAGCGAGATGACTGTGAGCGGCTGATGAGCATCAACTGCGAC 597
Db 461 ACTGTGTGATGTGAGAGCACAGATGACCTCTGCGCTGTGACGCGCTGCAACGAGTGA 520
Qy 598 TGGCTGAGCGCGCGAGAGGCGCATGTGACAGCGGCGGTACTGCGCTTGGGCTGACCC 657
Db 521 TGGTGTCCGCGGACGAGGCGCGGTGTACACCGGCGTCTGAGAGGCTTTCGCTTACTCCG 580
Qy 658 AGGCAAGATCAATAGTCTTTTACTGTGCTGCGCTTCTGCGCTGGGCGGAGAGGCGCA 717
Db 581 ACAAGAGCGCGCGCTGCGCTGCGCGCGCTTTCGACACGAGCGGTGTGCGGTGTCAGA 640
Qy 718 ACCTGCACTGTGAGTGGACCGCTGCGCGCGCTTCTGCGACATCAAGAGCTTACTGCG 777
Db 641 ACCTTCGCGGTATAGCGCGCGCGCTCTCCCGAATGATTCGCGAACCGCGGCGTGG 700
Qy 778 AGGACCGGCTCTGAGACGAGATGCGCTCTTGTGCGATGACCCGAGTGTGCGCTGATTCG 837
Db 701 GCGCGCGCATGTGAGACCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTACT 760
Qy 838 CCGGCGATGTTCCCGAGGCTGTC 860
Db 761 ACGGCAAGTCCCGAAGGCTTC 783

```

```

RESULT 15
US-10-156-761-1/c
; Sequence 1, Application US/10156761

```





FT Modified-site 532 /note= "potential N-glycosylation site"

XX MO9719177-A1.

XX 29-MAY-1997.

XX 22-NOV-1996; 96MO-AU00747.

XX 23-NOV-1995; 95AU-0006748.

XX (WOMEN-) WOMEN'S & CHILDREN'S HOSPITAL.

XX Anson DS, Blanch L, Hopwood JT, Scott H, Weber B;

XX WPI; 1997-298114/27.

DR N-PSDB; AAT67163, AAT67164.

XX Nucleic acid encoding mammalian alpha-N-acetylglucosaminidase -

PT used for the diagnosis and treatment of mucopolysaccharidosis type

PT IIB, also used in gene therapy

XX Claim 11; Page 46-50; 79pp; English.

XX Human alpha-N-acetylglucosaminidase (AA18017) can hydrolyse the

CC terminal alpha-N-acetylglucosamine residues at the non-reducing

CC terminus of fragments of heparan sulphate and heparin. Its amino

CC acid sequence was deduced from a human peripheral blood leukocyte

CC cDNA clone (AAT67163). Recombinant enzyme can be expressed in host

CC (pref. CHO) cells and may be expressed as a fusion to e.g. an

CC enzyme, reporter molecule, purification tag and/or signal sequence.

CC It can be used to treat alpha-N-acetylglucosaminidase deficiency,

CC for example in patients suffering from mucopolysaccharidosis type

CC IIB. Administration is by oral, i.v., i.p., enzyme replacement

CC therapy, gene therapy or other routes.

XX Sequence 743 AA;

XX

Query Match 100.0%; Score 3939; DB 18; Length 743;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAAVAAAVGVLLLAGAGAGADDEARAAVAVLALVRLGPPADFSVSEVERLAAKP 60

DB 1 MEAAVAAAVGVLLLAGAGAGADDEARAAVAVLALVRLGPPADFSVSEVERLAAKP 60

QY 61 GLDTYSLGGGGAARVRVRSSTGVAAAAGLHRYLRDPGCGHVAWSGQLRPLPVPAGE 120

DB 61 GLDTYSLGGGGAARVRVRSSTGVAAAAGLHRYLRDPGCGHVAWSGQLRPLPVPAGE 120

QY 121 LTEATPRRYRYQNVCTOSYFVWMDMARWEREIDMWALNGINIALAWSGOAIWQRYL 180

DB 121 LTEATPRRYRYQNVCTOSYFVWMDMARWEREIDMWALNGINIALAWSGOAIWQRYL 180

QY 181 ALGLTQAEINEFTGPAPLAWGMGNLHTWDGFLPSPWHIKOYLQHRVLDQMSFGMT 240

DB 181 ALGLTQAEINEFTGPAPLAWGMGNLHTWDGFLPSPWHIKOYLQHRVLDQMSFGMT 240

QY 241 VLPAFGHVEATRYFPQNVNTKMGWGHFNCSYCSFLIAPEDIFPISLFLRELI 300

DB 241 VLPAFGHVEATRYFPQNVNTKMGWGHFNCSYCSFLIAPEDIFPISLFLRELI 300

QY 301 KERGTDIYAGDTFENWQPPSSPSYLAATTAIVYEMTAVDTEAVMLQGMVLFQHPQF 360

DB 301 KERGTDIYAGDTFENWQPPSSPSYLAATTAIVYEMTAVDTEAVMLQGMVLFQHPQF 360

QY 361 WGAQIRAVLGAVERGLVLDLFASSQPYRTTASFOQPFVPCWLNHFGNHLFGAL 420

DB 361 WGAQIRAVLGAVERGLVLDLFASSQPYRTTASFOQPFVPCWLNHFGNHLFGAL 420

QY 421 EAVNGGEARLFPNSTMTVGMAPEGISQNEVYSLMELGWRKDPVDLAAWTSFAA 480

DB 421 EAVNGGEARLFPNSTMTVGMAPEGISQNEVYSLMELGWRKDPVDLAAWTSFAA 480

QY 481 RRYGVSHPDAGAAWRLLLRVYVNCSEACRGNHNSPLVRRPSIQMNTSIWNSDVEAW 540

DB 481 RRYGVSHPDAGAAWRLLLRVYVNCSEACRGNHNSPLVRRPSIQMNTSIWNSDVEAW 540

QY 541 RLILTSAPSLATSPAFRYDLDTROAVQGLVSLYTEAASATLSKLAALLAGCYL 600

DB 541 RLILTSAPSLATSPAFRYDLDTROAVQGLVSLYTEAASATLSKLAALLAGCYL 600

QY 601 ELPLALDEVILASDRFLLGSLWLBQARAAVSEADPEQNSRYQLTMGPEGNILDYAN 660

DB 601 ELPLALDEVILASDRFLLGSLWLBQARAAVSEADPEQNSRYQLTMGPEGNILDYAN 660

QY 661 KOLAGLVANYTTPRWRFLBALVDVAQGIIPFOQHGFDPKAVFOLECAFVLSKORYPQPR 720

DB 661 KOLAGLVANYTTPRWRFLBALVDVAQGIIPFOQHGFDPKAVFOLECAFVLSKORYPQPR 720

QY 721 GDTVDLAKKIFLKYPGPMVAGSW 743

DB 721 GDTVDLAKKIFLKYPGPMVAGSW 743

RESULT 2

ABB68119

ID ABB68119 standard; Protein; 690 AA.

XX ABB68119;

AC 26-MAR-2002 (first entry)

XX 26-MAR-2002 (first entry)

DT Drosophila melanogaster polypeptide SEQ ID NO 31149.

DE Drosophila melanogaster polypeptide SEQ ID NO 31149.

XX Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX WO200171042-A2.

PN 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2001; 2001WO-US09231.

PF 23-MAR-2001; 2000US-191637P.

XX 23-MAR-2001; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX 11-JUL-2000; 2000US-0614150.

PA (PEPE ) PE CORP NY.

XX (PEPE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

XX WPI; 2001-656860/75.

DR N-PSDB; ABL12222.

XX N-PSDB; ABL12222.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

PS Disclosure; SEQ ID NO 31149; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL10511), expressed DNA

CC sequences (ABL101840-ABL16175) and the encoded proteins

CC (ABBS5737-ABBS72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 690 AA;

XX



QY 674 RWRFLFALVDVSAQGFPOOHQFDKN 700  
 DB 61 RASKYFAVLTRELQNRSPGLEWRKD 87

RESULT 4  
 AAE05171  
 ID AAE05171 standard; Protein; 497 AA.  
 AC AAE05171;  
 XX  
 XX 12-SEP-2001 (first entry)  
 DT  
 DE Human drug metabolising enzyme (DME-2) protein.  
 XX  
 XX Human; drug metabolising enzyme; DME-2; immunosuppressive; gene therapy;  
 KW cytostatic; autoimmune disorder; inflammatory disorder; atherosclerosis;  
 KW osteoporosis; eye disorder; hepatic tumour; Addison's disease; cretinism;  
 KW rheumatoid arthritis; acquired immune deficiency syndrome; AIDS; anaemia;  
 KW developmental disorder; endocrine disorder; iritis; acromegaly; epilepsy;  
 KW thyrotoxicosis; pancreatic disorder; diabetes mellitus; obesity; adenoma;  
 KW gastrointestinal disorder; nodular hyperplasia; conjunctivitis; glaucoma;  
 KW actinic keratosis; metabolic disorder; dysphagia; anorexia; carcinoma;  
 XX cell proliferative disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..25  
 FT /label= Signal\_peptide  
 FT Protein 26..497  
 FT /note= "Mature drug metabolising enzyme (DME-2) protein"  
 FT Domain 1..18  
 FT /label= Transmembrane\_domain  
 XX  
 XX WO200151638-A2.  
 XX  
 XX 19-JUL-2001.  
 PD  
 PF 12-JAN-2001; 2001WO-US01174.  
 XX  
 XX 14-JAN-2000; 2000US-0176139.  
 PR 21-JAN-2000; 2000US-0177443.  
 PR 28-JAN-2000; 2000US-0178574.  
 XX  
 XX (INCYTE GENOMICS INC.  
 PA  
 PI Yang J, Baughn ME, Burford N, Au-Young J, Lu DM, Reddy R;  
 PI Ring HZ, Hillman JL, Yue H, Azimzal Y, Yao MG, Gandhi AR;  
 PI Nguyen DB, Tang YT, Lai P, Bandman O;  
 XX  
 DR WPI: 2001-425874/45.  
 DR N-PSDB; AAD09937.  
 XX  
 PT Drug metabolizing enzymes and encoding polynucleotides, useful for  
 PT diagnosing, treating and/or preventing autoimmune, inflammatory, cell  
 PT proliferative, developmental, endocrine, eye, metabolic, and  
 PT gastrointestinal disorders -  
 XX  
 PS Claim 1; Page 136-137; 133pp; English.  
 XX  
 CC The present sequence is human drug metabolising enzyme (DME-2) protein.  
 CC Human DME and its nucleic acid molecule are useful for the diagnosis,  
 CC treatment and prevention of disorders associated with increased or  
 CC decreased expression of DME. Examples of such disorders include,  
 CC autoimmune/inflammatory disorder such as acquired immune deficiency  
 CC syndrome (AIDS), rheumatoid arthritis, osteoporosis; cell proliferative  
 CC disorder such as actinic keratosis, atherosclerosis; developmental  
 CC disorder such as epilepsy, anaemia; endocrine disorder such as  
 CC acromegaly, cretinism, thyrotoxicosis; pancreatic disorder such as  
 CC diabetes mellitus; eye disorder such as conjunctivitis, glaucoma, iritis;  
 CC metabolic disorder such as Addison's disease, obesity; gastrointestinal  
 CC disorder such as anorexia, dysphagia and hepatic tumours including

CC nodular hyperplasia, adenomas and carcinomas. DME DNA is useful for  
 CC creating 'knockin' humanised animals (pigs) or transgenic animals (mice  
 CC or rats) to model human disease. DME DNA is also in useful is gene  
 CC therapy. DME and its immunogenic fragments are useful for screening  
 CC libraries of compounds in several drug screening assays.  
 CC  
 XX  
 SQ Sequence 497 AA:  
 QY  
 DB 345 AVNLL-----QGTFPHQGFQWGPQIRAVLGVPPGRLLVLDLFAESQPVYTRRTAS 396  
 14 AIFLLVLDLMDHRRQRRAARYPP--GLPLPLGL-----GNLLHYD----- 50  
 QY 397 FQCGPFTWCLNHFNGN-----HQLFGALRA-VNCGPEARL--FPNSTWV 439  
 51 FQNTFYCFQGLRRRPFQDVSLQAWTPVYVINGLAARALVTHGEDTADRPVPITQIL 110  
 QY 440 GTMAPEGT-----SQNEVYSLMAELGMRCDVPDILAWTSTPAARRYGVSHPD 489  
 111 GFPRSQGVFLARYGPAMREQRFRFSVSTLRNLGLK--KSLQWVTEEA----- 158  
 QY 490 AGAAMRLLRSVYNGSGACRGHNRSPILVRPSLSQWNTSIWNRSPVFEAMRLLITSAPS 549  
 159 -----CLCAAFANHSGRPF--RNGLLDKAV-----SNVI-----AS 188  
 QY 550 LATSPAFRYD-----LLDLTROAVQELVSLYEARSAYLSKEILASLIRAGCVLAF-- 601  
 DB 189 LTCGRFFEDDPFRLLDLQACGLKEBSGFLAEVINAAPVPHIPAL--AGVLFQFORA 246  
 QY 602 LPLALDEVL-----ASDSRFLIGSWLBEQARAAYSEADPYEONSRYQLTLMGPE 652  
 DB 247 FLTQDLBELTHERMTDPAQPPRDLTEAFILAK--KEKAKSPSSSFNDENLRIV----- 299  
 QY 653 GNILDYANKQLAGLVANYTTPRWRFLFLEALVDSVAQGFPOOHQFDKNVFOLEQAFVLAK 712  
 DB 300 GNLF-----LAGWVTTSTLLAWALLMLTHPVGQRV--QOEIDVIGQVHPHMAQO 350  
 QY 713 QRYP 716  
 DB 351 AHMP 354

RESULT 5  
 AAU33611  
 ID AAU33611 standard; Protein; 4342 AA.  
 XX  
 AC AAU33611;  
 XX  
 XX 14-FEB-2002 (first entry)  
 DT  
 DE Pseudomonas aeruginosa cellular proliferation protein #55.  
 XX  
 XX Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Pseudomonas aeruginosa.  
 XX  
 XX WO200170955-A2.  
 XX  
 XX 27-SEP-2001.  
 PD  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 XX 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.

XX (ELIT) ELITRA PHARM INC.  
 PA Hasebeck R, Ohlsen KU, Zyekind JW, Wall D, Trawick JD, Carr GU;  
 PI Yamamoto RT, Xu HH;  
 XX WPI: 2001-611495/70.  
 DR N-PSDB; AAS51470.  
 XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 PS Example 3; Seq ID No 5107; 511pp; English.  
 XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 4342 AA;  
 Query Match 3.1%; Score 121; DB 22; Length 4342;  
 Best Local Similarity 21.4%; Pred. No. 2.5;  
 Matches 178; Conservative 101; Mismatches 318; Indels 236; Gaps 44;  
 QY 4 YAVAA-----AVGVLLIAGGCA-----AGCEBARAAVRA-LVARKL 40  
 DB 1181 VAIQAERSPOLLVGLAIVKAGVYPLDPDYSERLAYMLAAGVELLITQAHLERL- 1239  
 QY 41 GPG-----PAADFESVERALAAKPGDVTSLGGGGAARVVGSGVAAGLHRYLRD 95  
 DB 1240 -PGEAGVTPLCDBSLKIDNWPQAGLHL--GDLNAVYVYTSGSTQPKGVG----- 1289  
 QY 96 FCGCHVAMSGSQLRPLRPPLPAVPGE--LTBATPNRYRYONVCTQSYSPVMDMARERE 153  
 DB 1290 --NTHALIA-ERLQWQATYTLDDGDVLMQKAP-----VSFVSVYV--E 1328  
 QY 154 IDWVALNGIMLALAMSGO-----EAIWQRYVALGLTGAINEFFTGPAFLANGRWG 205  
 DB 1329 CFWPLVIGCRIVLAAPGHRDPAELVLRQFGTTTHFVPLILQLFIDEEGVAAGCSLR 1388  
 QY 206 NLHTWMDPLPSPWHIKQLYLQHRVLDQMSFGMTFVLPAPAGVPEAVTRVFPQVNVYTK 265  
 DB 1389 RLFCGGEALPAE-----LRNRYLQR-----LPAVALHNRVGPET--AINVT-- 1428  
 QY 266 GSWHFPNCSYSGSLFAPD-----PIPIIGSLFLRLILIEFGTDHLYGADTFNEM---- 317  
 DB 1429 ---HMOC-----RAEDGERSPIGRPLGNVVCRLDAEF--NLLPAGVAGELCTIG 1473  
 QY 318 -----OPSPSEPSYLAATAVVEAMTAVDTEAVMLLOGL-----FCHQPOFWG- 362  
 DB 1474 IGLARVIGRPAJSAERFVADPFSAGGERLYRTGDRARNADGVLEVLGRLDQVVLRRG 1533  
 QY 363 ---PAQIRAYLGAVP---RGRLLVLDLPAESQPV--YTRTASQGGPFIWMLHNGGNH 414  
 DB 1534 RIEPEEIQAARLLAOPGVAQAVVIREGVAQSOLGVYTGAVGAEA-----EQHQ 1584  
 QY 415 GLFALBAVNGGPEARLFPNSTWVGTMAGPEGISQEVVYSIMAEIGWRK---DPVDD 470

DB 1585 RLRAALQA-----ELPEYVWPTQLMRLAQMPLGPS-GILDTRALPEPWOQREHVEPTE 1638  
 QY 471 L-----AAWVTSFAARRYG-----YSHPDAGAMPLLRSYVNC-- 505  
 DB 1639 LQRRALAIWSEVLGIPRVGLDDDFELQGHSLIARIVSRTRQACDVLPRLALPEASEL 1698  
 QY 506 -----GEACRGHNRSPVLRPSLQMTSIIWYNSRDVEAMRLLITSAPLSATSPAF 556  
 DB 1699 EAPCEQVBAQAAGRTDGHGAIIRIDREQPVPLSYQQQRMFLMQI-----EPD--SEAY 1751  
 QY 557 RY-----DLIDLR--QAVQELVSLYEARSAYLSKELASLRA---GSV-LAYELL 603  
 DB 1752 NVGILARISGLPIDVAFPAALQALVQ-RHETLRTFSPVDGVPPVORVHGDGLHMDQDF 1810  
 QY 604 PALD-----EVLASDSRFLGSMLEQA---RAAAVSEADPEYQNSRYQLTWGPE 652  
 DB 1811 SALDSDSRQOHLQTLA-DSEARFPEDLESGLPLRVCMKABREHYLVVTLHHIV---TE 1866  
 QY 653 GNILDYANKQLAGVANYTTRMRFLFALVDVAGGIFPFGQHPDKVPCLE 705  
 DB 1867 GWAMDIPARELGALYEAPLDR-----ESPLEPLVQ--YIDYSVWQRE 1908  
 RESULT 6  
 ID AAM22601 standard; Protein; 4472 AA.  
 XX AAM22601;  
 AC AAM22601;  
 XX 27-FEB-1998 (first entry)  
 DT  
 XX  
 DE Ty lactone synthase ORF1 protein.  
 XX Ty lactone synthase gene cluster; tylG gene; multifunctional protein;  
 KW polyketide; ty lactone synthesis; antibiotic; tylostin.  
 XX Streptomycetes fradiae.  
 XX  
 FT Key Location/Qualifiers  
 FT Misc-difference 1  
 FT /note= "encoded by GTG"  
 FT Domain 43..447  
 FT /note= "ketosynthase domain, KSq(s)"  
 FT Domain 586..914  
 FT /note= "acyltransferase domain, At(s)"  
 FT Domain 954..1038  
 FT /note= "acyl carrier protein domain, ACP(s)"  
 FT Domain 1060..1483  
 FT /note= "ketosynthase domain, KS1"  
 FT Domain 1606..1934  
 FT /note= "acyltransferase domain, Att1"  
 FT Domain 2199..2378  
 FT /note= "ketoreductase domain, KR1"  
 FT Domain 2469..2552  
 FT /note= "acyl carrier protein domain, ACP1"  
 FT Domain 2576..2999  
 FT /note= "ketosynthase domain, KS2"  
 FT Domain 3149..3477  
 FT /note= "acyltransferase domain, Att2"  
 FT Domain 3502..3687  
 FT /note= "denitratase domain, DN2"  
 FT Domain 4016..4200  
 FT /note= "ketoreductase domain, KR2"  
 FT Domain 4302..4385  
 FT /note= "acyl carrier protein domain, ACP2"  
 PN EP791655-A2.  
 XX 27-AUG-1997.  
 PD 19-FEB-1997; 97EP-0301056.  
 XX



CC gene cluster of the invention. The gene cluster is also referred to as  
 CC the *tylG* gene, and was isolated from *Streptomyces fradiae*. These  
 CC sequences are multifunctional proteins which direct the synthesis of the  
 CC polyketide tetraketone, isolated from *Streptomyces fradiae*. Tetraketone is  
 CC the basic building block of the antibiotic tylosin. The DNA sequence can  
 CC be modified so as to alter the type of carboxylic acids incorporated,  
 CC the number of carboxylic acids incorporated and/or the post-condensation  
 CC reactions performed, thereby resulting in novel tylosin-related  
 CC polyketides.

XX Sequence 3729 AA;

Query Match 3.0%; Score 120; DB 18; Length 3729;

Best Local Similarity 22.4%; Pred. No. 2.5;  
 Matches 192; Conservative 72; Mismatches 265; Indels 328; Gaps 47;

4 VAVAAV---GVLLLAGAGAGADEAREAAVAVLVARLLG-----PGPAD 47  
 668 VDVAAVGHSGEILAAITVAGALSLE--DAAVVALPAIGLYLAGRMAAVPLPAGE 725  
 48 FSVSVERALAAKPGLDITYSLGGGAA---RVRVRSSTGVAAAAGLI----- 90  
 726 ---VEAGLAKMPGEVAAVAVGPASTVVGDRRAVAGYVAVCAEGVQARLI PVDYASHS 781  
 91 RYLARDPGC-CHVAMSGSGLRPR-ELPRAV-----PGELEATNRRRYQNCTQSYSPV 143  
 782 RHVEDLKGELERVLISGIRPSRPVPCSTVAGQPGEPV----- 820  
 144 WMDWARWER---EIDMMALNGINL-----ALAMSGEALIMORVYLA 181  
 821 -FDAGVWFRNLRNVERSAVVGLEEGHRRFIEVSAHPVLAIBGTAA--DSVVA 877  
 182 LGLTQAE---INEFTGPAPLAMGRMGNLTWDGFLPSPHAIKQLYL---QHRVLDQM 233  
 878 TGLTRRDDSPHRLLTSTA-EAMAHGATL-TWDPALPPG-HLTTLPTYPENHHWYLD-- 932  
 234 RSFGMTVLPAF-----AGHVEAVT-RVPPQVNTKM-GSWGHNCSYSCS 278  
 933 ---TTTTRATTTQSTPDANPADALPYKVSWMRLDQDSTLRADGR-- 978  
 279 FLAPEDPIPTIGSLFLRELIR-----EFGTDH-----IYGADTNEMOPSS 322  
 979 LTVVPEASADPSVAEGVARELTARGATVESLTVEPGADRRLRGLVDAIEREAGPLNG 1038  
 323 EBSVLA-----AATTAVYEMTAVDTEA-WMLQGLFPHQPFMG 362  
 1039 TVSLALAGADGADGARPVPVAGLAASLALICAGAGTEAGIM----- 1083  
 363 PAQIRAVLGAVERGLLVLDFAPESQPV---YRTASFOGQPTWMLNFGNGHGLFG- 418  
 1084 -AVTRGAVVAAPG---DVPAPSQALMVGFRVAGIE-LPHCWGGLDLPLTGPDPGSP 1135  
 419 ---ALEAVNGPEEARLFPNSTWMTGT-GMAPEGISONEVVYSIMAEI 461  
 1136 RQLAATLAGRAEDQVALRA--SGAYGRRLVRASAGAGADGMRPG-----TVLVV 1184  
 462 GMRKDPVVDLAAMTSPAARY---GVSHP---DAGAAWRLLSRYNCSGACRGHNR 514  
 1185 GDTNEVAGPLVRWLLGNGARRVTLGSLGPLBEELAVARVY-----APCOPADR 1236  
 515 SPLVRPSLQMTSITWNRSDVPEAMRLLTSAPLSATSPAFRYDLIDLTRQA----- 567  
 1237 PALRTLLAEQAPTV-----LVAPPAVPTP-----LAEMTAELATALSA 1277  
 568 ---VOELSLVYE-----EARSAYLSKELASLL 592  
 1278 KTGIVDLDSLDDPDLBEGELDAFVBSVAGWGAGGGYAGTAYLD-ALAECK 1336  
 593 RAGG-VIAVEELPALDEVLASDSRFTLGSWLEQA-----PAAV 630  
 1337 RAGGILPTSVAMTWLG-TPAADS--LGEQMSPAGITPLDPAASLDALARAVERAGCV 1392  
 631 SEAEADFEONSRYOLT 647

DB 1393 TVADIDMERFASAYTAT 1409

RESULT 8  
 AAB30570  
 ID AAB30570 standard; Protein; 871 AA.  
 XX  
 XX AAB30570;  
 DT 19-MAR-2001 (first entry)

DE A splice variant of a signal transduction polypeptide.

XX Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;  
 XX congestive heart failure; dilated congestive cardiomyopathy;  
 XX hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
 XX mitral valve disease; aortic valve disease; tricuspid valve disease;  
 XX myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
 XX atherosclerosis; cardiac tumour; microbial infection; splice variant.

XX Homo sapiens.

XX WO200063381-A1.

XX 26-OCT-2000.

XX 11-APR-2000; 2000WO-US09488.

XX 16-APR-1999; 98US-0129553.

XX (SCIO-) SCIOS INC.

XX Zeng W, Stanton L, Kong H;

XX WPI; 2001-007013/01.

XX Novel H19G5 polypeptides capable of regulating signal transduction and  
 XX exhibiting kinase activity useful for identifying antibodies to treat  
 XX cardiac diseases, and additional mediators of signal transduction  
 XX  
 XX Claim 1; Page 74-76; 81pp; English.

XX The present sequence represents a splice variant of human in signal  
 XX transduction polypeptide. The polypeptide is designated H19G5. The  
 XX protein is capable of regulating signal transduction and exhibits kinase  
 XX activity. The H19G5 transcript is expressed in the heart. H19G5  
 XX polypeptides and polynucleotides are useful for preventing or treating a  
 XX cardiac disease, such as congestive heart failure, dilated congestive  
 XX cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,  
 XX mitral valve disease, aortic valve disease or tricuspid valve disease,  
 XX angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,  
 XX arterial or rheumatic hypertension, arteriosclerosis, atherosclerosis  
 XX and cardiac tumours in humans. The polypeptide is also useful for  
 XX detecting the expression of a protein capable of regulating signal  
 XX transduction or the expression of a protein capable of acting as a donor  
 XX or acceptor molecule of a phosphate group. The monoclonal antibodies can  
 XX be used as probes for detecting discrete antigens expressed by tissue or  
 XX cell samples, and therefore used in humans for localization and  
 XX monitoring of microbial infection.

XX Sequence 871 AA;

Query Match 3.0%; Score 118.5; DB 22; Length 871;

Best Local Similarity 18.6%; Pred. No. 0.34;  
 Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

19 GGAAGDEAREAAVAVLVARL-LGSPAD---FSVSRERLAAKPGD-----TY 65  
 21 GSASQASSQVSSLVGSSQVTEPGSLDABGWQAEADSDSTPLQRPQEQATMKF 80  
 66 SLGG-GGAARVVRSGSTVAAAAGLHRYLRDFCG-----CHVAMSGSLR----- 109

XX		cadiovascular disease; hypertension; vasculitis; myocarditis; obesity;
KW		congestive heart failure; ischemic heart disease; lung tumour; gout;
KM		fatty liver; Niemann-Pick's disease; gene therapy.
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Domain	575..827
FT		/note= "Eukaryotic protein Kinase domain"
FT	Domain	580..812
FT		/label= Protein_Kinase_domain
PN		
XX		
PD		
XX		
PE	20-DEC-2001.	
XX		
PR	14-JUN-2001; 2001WO-US19444.	
XX		
PR	15-JUN-2000; 2000US-212073P.	
PR	23-JUN-2000; 2000US-21367P.	
PR	30-JUN-2000; 2000US-215651P.	
PR	07-JUL-2000; 2000US-21605P.	
PR	13-JUL-2000; 2000US-218372P.	
PR	25-AUG-2000; 2000US-228056P.	
XX		
PA	(INCY-) INCYTE GENOMICS INC.	
PI	Yue H, Lal P, Bandman O, Borowsky ML, Au-Young J, Lu Y;	
PI	Gandhi AR, Tripathy CM, Wala NK, Yao MG, Lu DM, Greenwald SR;	
PI	Ramkumar J, Griffin JR, Kearney L, Burford N, Nguyen DB, Tang YT;	
PI	Baughn MR, He A, Thornton M, Hafalia A, Patterson C, Gurrarjan R;	
PI	Lo P, Khan F, Reipon SA, Azimzai Y, Policky JL, Ding L,	
PI	Grether M, Elliott VS, Thangavelu K, Batra S, Ison CH;	
XX		
DR	WPI; 2002-090207/12.	
DR	N-PSDB; AAD26467.	
PT		
PT	New polypeptides, useful for diagnosing, treating or preventing	
PT	diseases of growth and development, cardiovascular and lipid, and	
XX	diseases such as cancer, comprise human kinase polypeptides -	
XX		
PS	Claim 1; Page 164-165; 197pp; English.	
XX		
CC	The invention relates to human kinase PKIN proteins and their	
CC	corresponding cDNAs. A composition containing PKIN agonist is useful for	
CC	treating a disease or condition associated with decreased expression of	
CC	PKIN and a composition comprising PKIN antagonist is useful for treating	
CC	a disease or condition associated with overexpression of PKIN. The	
CC	disorders include cancer (leukaemia, adenocarcinoma, lymphoma, melanoma,	
CC	myeloma, sarcoma, leiomyosarcoma, Hodgkin's disease), immune disorder	
CC	(Acquired Immune Deficiency Syndrome (AIDS)), asthma, Addison's disease,	
CC	atherosclerosis, anaemia, allergies, adult respiratory distress syndrome,	
CC	autoimmune thyroiditis, gout, bronchitis, Crohn's disease, diabetes	
CC	mellitus, multiple sclerosis, pancreatitis, psoriasis, Reiter's syndrome,	
CC	osteoarthritis, osteoporosis, pancytopenia, uveitis, ulcerative colitis,	
CC	rheumatoid arthritis, Sjogren's syndrome, warts, viral, protozoal and helminthic infections)	
CC	bacterial, parasitic, fungal, viral, protozoal and helminthic infections)	
CC	growth and development disorders (arteriosclerosis, cirrhosis, hepatitis,	
CC	Cushing's syndrome, hypothyroidism, cerebral palsy, cataracts); cardio	
CC	vascular disease (arteriovenous fistula, hypertension, vasculitis,	
CC	aneurysms, congestive heart failure, angina pectoris, myocarditis,	
CC	ischemic heart disease, chronic bronchitis, lung tumours); lipid	
CC	disorder (fatty liver, Fabry's disease, Niemann-Pick's disease,	
CC	hypercholesterolaemia, obesity). PKIN DNA is useful for assessing	
CC	toxicity of a test compound and in gene therapy. The present sequence	
CC	is human PKIN-20 protein.	
XX		
SQ	Sequence 871 AA;	
XX		
Query Match	3.0%; Score 118.5; DB 23; Length 871;	
Best Local Similarity	18.6%; Pred. No. 0.34;	
Matches .165; Conservative	97; Mismatches 307; Indels 319; Gaps 40;	

```

QY 19 GGAAGDEAREAAVALVARL-LGPGPAD---FSVSVERALLAAKGLD-----TY 65
D 21 GSASQSSSSQVSLRVSSQVGTPEPSLDAGMTQEAEDLSSTPTLQRPQVMTKRF 80
QY 66 SLGG-GGAARVRVRGSGVAAAGLHRYLDFCG-----CHVANGSGQLR-----109
D 81 SLGGRGYAGVAGYGTAFAGDAG-----GMLGGGPMARLMAVSSQSEEEQEARAE 134
QY 110 -----LP---RPLPAVPGELTEATPNRRYYQNVCTQSYSFVMMQWAME 151
D 135 SQSEEQEARAEPLPQVSARPVDEVGAPTRSSP-----EPTPWE 175
QY 152 -----REI---DMALNGINLAL-----AMSGEAIWQRYIAL-----182
D 176 DIGQVSLVQIRDLISGDAEADTISLDISEVDPAYLMSLDYIKYLPFEEMTFRKVPKSA 235
QY 183 -----GLTQAEINEF-----FTGP-----AFLAMGRGNLHTWGP 213
D 236 QPEPSPMAEEELAEFPPTWMPGELGPHAGLETBESDVALLAEAAVGRKRMSSP 295
QY 214 LPPSWHIKQLYLQHRVLDQMSFGMTFVLPAFAGHV-----PEAVTRVFPQVNTKMG 266
D 296 SRSIFHPG---RHLPLDEPAELGLRRVYASVEHSIRILKGRPEGLEKEGP---RKTG 349
QY 267 SWGHFNCSYCSFILAEDEPIFPIISLFLRELKERGDHITGADTFNEMQPSSESPY 326
D 350 GLASFRLSGKXMDRAP-----TFUREL-----SDETTV 378
QY 327 LAAATTAAYEAMTAVIDEAVMLQGMFLQHQPOFWGPAQIRAVLAGVPRG---RLVLVDLF 384
D 379 LGQSVTLACQVSAQPAQATMSXGABLE-----SSSRVLISATLKNFOLLTLIVVV 430
QY 385 AESRPVYTRTASFGQGFIMCMHFGGNGGLFGALEAVNGGSEARLFPNGTMTGMA 444
D 431 AEDIGVYI-----CSVSNALGYITTTGVLR-----KAER-PPSS-----463
QY 445 PEGISQNEVVYSLMAELMRKDPVPD-----LAAW---VTSFAARRYGSHPDAGAA 493
D 464 -----PCPDIGEYVADGVLLVWKPVSXGPTVYIVOCSSLEGSS 501
QY 494 WRLLRSVYNC-----SGEAC 509
D 502 WTLASDIFPCVLTSLKLSRGTYTPTACVSKAGMPYSSPEBOVLLGGPSHAEBS 561
QY 510 RGHNRSLVVRPSLQWMTSIWYNR-SVDFEAMRLLLTSAPSLATSPAFRDLDLTR---Q 566
D 562 QGRSAQFLPSTKTAFTQIQGRFSYVRQWEKASGRALAAKIIPHDKTAIVREYE 621
QY 567 AVQELVSLYVEARSAVLS-KEIASLLRAGGVLAPELLPALDEVLASDSRFLGSMLEQA 625
D 622 ALKGLRHPHLAQHLAAVLSPRHLVILIEL--CSGPEILLPCIAE-RASYSESEVKDYLMQW 678
QY 626 RAAA-----VSEADAYEON---SRQYLTIMGEGNILLYANKQLAGLVANYTTRML 677
D 679 LSAIYVHNGHITLHLDJRSENMITEYNLL-----KVDLGNAGSISQEKVLPSPDKFO 732
QY 678 FLEALVDSVAQG---IPFOQHOFKXVFOI-EQAFTVLSKQRYPSOPRG 721
D 733 YLEMAWELLEGGQAVP-----QTDIMALGVAFIMLSAEYVSSSG 774

```

```

KW hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
KW mitral valve disease; aortic valve disease; tricuspid valve disease;
KW myocardial infarction; cardiac arrhythmia; arteriosclerosis;
KW atherosclerosis; cardiac tumour; microbial infection.
OS Homo sapiens.
PN WO20063381-A1.
PD 26-OCT-2000.
PF 11-APR-2000; 2000MO-US09488.
PR 16-APR-1999; 99US-0129553.
PA (SCIO-) SCIOS INC.
PI Zeng W, Stanton L, Kong H;
DR WPI: 2001-007013/01.
DR N-PSDB; AAC62285.
PT Novel h19G5 polypeptides capable of regulating signal transduction and
PT exhibiting kinase activity useful for identifying antibodies to treat
PT cardiac diseases, and additional mediators of signal transduction -
PS Claim 1; Page 55-57; 81pp; English.
XS
CC The present sequence represents a human protein with putative function
CC in signal transduction. The polypeptide is designated H19G5. The protein
CC is capable of regulating signal transduction and exhibits kinase
CC activity. The H19G5 transcript is expressed in the heart. H19G5
CC polypeptides and polynucleotides are useful for preventing or treating a
CC cardiac disease, such as congestive heart failure, dilated congestive
CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,
CC mitral valve disease, aortic valve disease or tricuspid valve disease,
CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,
CC and cardiac tumors in humans. The polypeptide is also useful for
CC detecting the expression of a protein capable of regulating signal
CC transduction or the expression of a protein capable of acting as a donor
CC or acceptor molecule of a phosphate group. The monoclonal antibodies can
CC be used as probes for detecting discrete antigens expressed by tissue or
CC cell samples, and therefore used in humans for localization and
CC monitoring of microbial infection.
SQ Sequence 1351 AA;
Query Match 3.0%; Score 118.5; DB 22; Length 1351;
Best Local Similarity 18.6%; Pred. No. 0.69;
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;
QY 19 GGAAGDEAREAAVALVARL-LGPGPAD---FSVSVERALLAAKGLD-----TY 65
D 21 GSASQSSSSQVSLRVSSQVGTPEPSLDAGMTQEAEDLSSTPTLQRPQVMTKRF 560
QY 66 SLGG-GGAARVRVRGSGVAAAGLHRYLDFCG-----CHVANGSGQLR-----109
D 81 SLGGRGYAGVAGYGTAFAGDAG-----GMLGGGPMARLMAVSSQSEEEQEARAE 614
QY 110 -----LP---RPLPAVPGELTEATPNRRYYQNVCTQSYSFVMMQWAME 151
D 135 SQSEEQEARAEPLPQVSARPVDEVGAPTRSSP-----EPTPWE 655
QY 152 -----REI---DMALNGINLAL-----AMSGEAIWQRYIAL-----182
D 176 DIGQVSLVQIRDLISGDAEADTISLDISEVDPAYLMSLDYIKYLPFEEMTFRKVPKSA 715
QY 183 -----GLTQAEINEF-----FTGP-----AFLAMGRGNLHTWGP 213
D 236 QPEPSPMAEEELAEFPPTWMPGELGPHAGLETBESDVALLAEAAVGRKRMSSP 775
QY 214 LPPSWHIKQLYLQHRVLDQMSFGMTFVLPAFAGHV-----PEAVTRVFPQVNTKMG 266

```

```

Db 776 SSSLFHPG--RHLPDEPAELGLRERKASVEHISRLKGRFGLKEGPP--RKKP 829
QY 267 SMGHFNCSYSCSFLIAPEDPIPIIGSLFLRELLEFGTDHIYADTFNEMOPSSSEPSY 326
Db 830 GLASFRLSGLKMWRAP-----TFRLREL-----SDETVV 858
QY 327 LAAATTAVTEAMTAVTEAVTWLQGLFQHPQFQWGPQAQIRAVLGAVPRG--RLVYLDLF 384
Db 859 LGQSVTLACQVSAQPAQAATWKSQDAPLE-----SSSRVLISATLKNFQLTLIVVV 910
QY 385 AESQPYVTRTASFOGQPFIMCMILNFGNGLFGALNAVNGGPEAARLFNPNSTWGTGMA 444
Db 911 AEDLGYVT-----CVSNALGVTYTTGVLR-----KAEV--PSSS----- 943
QY 445 PEGISQNEVYYSIMAEIGWRKDPVD-----LAAM--VTSFAARRGVSHPDAGAA 493
Db 944 -----PCPDIGEVADGVLLVWKPVESYGPVTYIVQCSLEGGS 981
QY 494 WRLLSRVNC-----SGEAC 509
Db 982 WTTLASDITFDCCYLTSKLSRGTYTFRTCVSKAGNGVSPSEQLGSHLASSEES 1041
QY 510 RGHNSPLVRRPSLQNNISIWYNR--SDVEAWRLLLTSAPLSATSPAFRYDLDLTR--Q 566
Db 1042 QGRSAQPLPSTKTFARFQIQGRGFEVVRQCEKASGRALAKIIPYHKQKTAVALREYE 1101
QY 567 ANQELVSLYEERKSYIS--KELASLIRAGVLAWEILPALDEVIASRPLFSWLEQA 625
Db 1102 AKGLHHPHQAQMAAYLSPRHVLILEL--CSGPELLPCLAE--RAISYSESEVKDYLMQM 1158
QY 626 RAAA-----VSEAEADFEYQN--SRVQLTLWGPBGNILIDYANKOLAGLVANYTPRWRL 677
Db 1159 LSAIQYLHNOHILHLDLRSENMITEYNLL-----KVYDLDGNAQSLQSEKVLPSDKFKD 1212
QY 678 FLEALYDVSAAQ---IPFOQHOPDKNVPOL--EQAFPLSKORPPSQPRG 721
Db 1213 YLETPAPELLEGGAVP-----QTDIMAGVTAFTMLSAEYFVSSEG 1254

RESULT 11
AAB30568
ID AAB30568 standard; Protein; 1610 AA.
AC AAB30568;
DT 19-MAR-2001 (first entry)
DE A full length human signal transduction polypeptide.
XX Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;
XX congestive heart failure; dilated congestive cardiomyopathy;
XX hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;
XX mitral valve disease; aortic valve disease; tricuspid valve disease;
XX myocardial infarction; cardiac arrhythmia; arteriosclerosis;
XX atherosclerosis; cardiac tumour; microbial infection.
OS Homo sapiens.
XX
XX WO200063381-A1.
XX
XX 26-OCT-2000.
XX
XX 11-APR-2000; 2000WO-US09488.
XX
XX 16-APR-1999; 99US-0129553.
XX
XX (SCIO-) SCIOS INC.
XX
XX Zeng W, Stanton L, Kong H;
XX
XX WPI; 2001-007013/01.
XX
XX N-PSDB; AAC62286.

```

```

XX The present sequence represents a human protein with putative function
PT Novel h19G5 polypeptides capable of regulating signal transduction and
PT exhibiting kinase activity useful for identifying antibodies to treat
PT cardiac diseases, and additional mediators of signal transduction
XX
XX Claim 1; Page 61-65; 81pp; English.
XX
XX The present sequence represents a human protein with putative function
in signal transduction. The polypeptide is designated H19G5. The protein
is capable of regulating signal transduction and exhibits kinase
activity. The H19G5 transcript is expressed in the heart. H19G5
polypeptides and polynucleotides are useful for preventing or treating a
cardiac disease, such as congestive heart failure, dilated congestive
cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,
mitral valve disease, aortic valve disease or tricuspid valve disease,
angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,
arterial or rheovascular hypertension, arteriosclerosis, atherosclerosis
and cardiac tumours in humans. The polypeptide is also useful for
detecting the expression of a protein capable of regulating signal
transduction or the expression of a protein capable of acting as a donor
or acceptor molecule of a phosphate group. The monoclonal antibodies can
be used as probes for detecting discrete antigens expressed by tissue or
cell samples, and therefore used in humans for localization and
monitoring of microbial infection.
SQ Sequence 1610 AA;
Query Match 3.0%; Score 118.5; DB 22; Length 1610;
Best Local Similarity 18.6%; Pred. No. 0.9;
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;
QY 19 GGAAGDEBARBAAVRALVRL--LQGPBPD--FSVSVERALAKGLD-----TY 65
Db 760 GSASQASSQVSSISLRVSSQVGTPEPGLDABEGMTQAEPLDSSTPLQRPQDQATWRKF 819
QY 66 SLIG--GGAARVRRGSGVAAAAGLHRYLDPG-----CHVWMSGQLR----- 109
Db 820 SLGGRGGVAGVAGYTAFGDAG-----GMLGQPMARIRIMAVSQQSEEGQEARAE 873
QY 110 -----LP--RPLPAVPEGLTEATPNRYRYQNVCTQSYFVWDMARWE 151
Db 874 SQSEEGQEARAEPLPVASRPVDEVRAPTRRSP-----EPTPWE 914
QY 152 -----RET--DMALNGINIAL-----AMSQGEALWQRYVAL 182
Db 915 DIGQVSLVQIRDSGDREARADTISLDISEVDPATLMSDYDKILPFRFMIRKPKSA 974
QY 183 -----GLTQAEINEF-----FTGP-----AFLWGMGNLHTWDGP 213
Db 975 QPEPSPMAEHELAEFEPTWPMPEGLGPHAGLEITESESDVALLAEAVGRKRKSSP 1034
QY 214 LPPEMHIKQLYLQHRVLDQKRSFGMTPLVLPAPAGV-----PEAVTRFPQVNTKMG 266
Db 1035 SRSLSFHPG--RHLPDEPAELGLRERKASVEHISRLKGRFGLKEGPP--RKKP 1088
QY 267 SMGHFNCSYSCSFLIAPEDPIPIIGSLFLRELLEFGTDHIYADTFNEMOPSSSEPSY 326
Db 1089 GLASFRLSGLKMWRAP-----TFRLREL-----SDETVV 1117
QY 327 LAAATTAVTEAMTAVTEAVTWLQGLFQHPQFQWGPQAQIRAVLGAVPRG--RLVYLDLF 384
Db 1118 LGQSVTLACQVSAQPAQAATWKSQDAPLE-----SSSRVLISATLKNFQLTLIVVV 1169
QY 385 AESQPYVTRTASFOGQPFIMCMILNFGNGLFGALNAVNGGPEAARLFNPNSTWGTGMA 444
Db 1170 AEDLGYVT-----CVSNALGVTYTTGVLR-----KAEV--PSSS----- 1202
QY 445 PEGISQNEVYYSIMAEIGWRKDPVD-----LAAM--VTSFAARRGVSHPDAGAA 493
Db 1203 -----PCPDIGEVADGVLLVWKPVESYGPVTYIVQCSLEGGS 1240
QY 494 WRLLSRVNC-----SGEAC 509

```



Qy	626	RAAA-----VSEADAFYQCN---SRVQTLTGWGPEGNTLIDYANKQGLVNAVYTPFMR	677
Db	1426	LSATQYVHNGHILHLRLRENNMILTEYNLL-----KVYDLGNQGSUQEKVLPSSDKFXD	147
Oy	678	FILEALYDSVAQG---IPFOGHQEDKNVFOI-EQAVLSKQRPQSPRG	721
Db	1480	YLETMAPELLGEGGAVP-----QTDIMAIQVTAIFMISAEPVSSSG	1521
RESULT 13			
AAE24151			
ID	AAE24151	standard; Protein; 1665 AA.	
AC	AAE24151;		
DT	23-SEP-2002	(first entry)	
XX			
DE	Human kinase (PKIN)-22 protein.		
XX			
KW	Human; kinase; PKIN; cancer; immune system disorder; atherosclerosis;		
KW	acquired immune deficiency syndrome; AIDS; Addison's disease; allergy;		
KW	asthma; multiple sclerosis; psoriasis; arteriosclerosis; cirrhosis;		
KW	development; hepatitis; cardiovascular; hypertension; drug screening;		
KW	myocardial infarction; Goodpasture's syndrome; lipid disorder; growth;		
KW	fatty liver; Gaucher's disease; Niemann-Pick's disease; anorectic;		
KW	hypercholesterolemia; obesity; gene therapy; cytostatic; anti-hiv;		
KW	neuroprotective; hepatotrophic; hypotensive; cardiac; nephrotrophic;		
KW	hyperlipidaemia; enzyme.		
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	Domain	68..118	
FT		/note= "Immunoglobulin domain"	
FT	Domain	165..418	
FT		/note= "Eukaryotic protein kinase domain"	
FT	Domain	167..401	
FT		/note= "Protein kinase domain"	
FT	Domain	1174..1235	
FT		/note= "Immunoglobulin domain"	
FT	Domain	1369..1621	
FT		/note= "Eukaryotic protein kinase domain"	
FT	Domain	1372..1606	
FT		/note= "Protein kinase domain"	
FN	WO200233099-A2.		
XX			
PD	25-APR-2002.		
XX			
PF	20-OCT-2001; 2001WO-US47728.		
XX			
PR	20-OCT-2000; 2000US-242410P.		
PR	27-OCT-2000; 2000US-244068P.		
PR	03-NOV-2000; 2000US-247808P.		
PR	09-NOV-2000; 2000US-247672P.		
PR	16-NOV-2000; 2000US-249565P.		
PR	22-NOV-2000; 2000US-252730P.		
PR	01-DEC-2000; 2000US-250807P.		
XX			
PA	(INCY- ) INCYTE GENOMICS INC.		
XX			
PI	Gururajan R, Baughn MR, Walla NK, Elliott VS, Xu Y, Arvizu C;		
PI	Yao MG, Ramkumar J, Ding L, Tang YT, Hafalia AA, Nguyen DB;		
PI	Gandhi AR, Lu Y, Yue H, Buford N, Bandman O, Tribouley CM;		
PI	Lai PG, Recipon SA, Lu DAM, Borowsky ML, Thornton M, Swarnaker A;		
PI	Thangavelu K, Khan FA, Ison CH;		
XX			
DR	WPI: 2002-454603/48.		
XX			
DR	N-PSDB: AAD38865.		
XX			
PT	New human kinase polypeptide, for diagnosing, preventing and treating		
PT	cancer, immune system disorders, growth and development disorders,		
PT	cardiovascular disorders and lipid disorders		

XX Claim 1; Page 182-186; 210pp; English.

CC The invention relates human kinases (PKIN) and their corresponding  
CC nucleic acid sequences. PKIN and its DNA are useful for diagnosing,  
CC treating and preventing cancer, an immune system disorder (e.g., AIDS),  
CC acquired immune deficiency syndrome (AIDS), Addison's disease, allergy,  
CC asthma, arteriosclerosis, multiple sclerosis, psoriasis, disorders  
CC affecting growth and development (e.g., arteriosclerosis, clefts,  
CC hepatitis), cardiovascular disorder (e.g., hypertension, myocardial  
CC infarction, Goodpasture's syndrome), and a lipid disorder (e.g., fatty  
CC liver, Gaucher's disease, Niemann-Pick's disease, hypercholesterolemia,  
CC hyperlipidemia, obesity), and for assessing the effects of exogenous  
CC compounds. Anti-PKIN antibody is useful in a diagnostic test for a  
CC condition or a disease associated with the expression of PKIN in a  
CC biological sample. A composition comprising PKIN or an agonist or  
CC antagonist of PKIN is useful for treating a disease or condition.  
CC associated with decreased or increased expression of functional PKIN.  
CC PKIN is useful in a number of drug screening techniques and to analyze  
CC the proteome of a tissue or cell type. PKIN DNA is useful for creating  
CC knockin humanised animals or transgenic animals to model human diseases,  
CC and in somatic or germline gene therapy. The present sequence is human  
CC PKIN protein.

CC

CC Sequence 1665 AA;

XX

SX

Query Match 3.0%; Score 118.5; DB 23; Length 1665;  
Best Local Similarity 18.6%; Pred. No. 0.95;  
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

Dy 19 CGAAGDARERAAVRALVARL-LGPGAAD---FSVSEVERALAKPLD-----TY 65

Dd 815 GSASQASSSSSVSLRVSQVTGPGLDAEGTGAEDLDSTPTLPORPQCVTKRF 874

Oy 66 SLGS-GGAARVRVGSTGTVAAGAHLHYLRDFCG--CHVASGSOLR-----109

Dd 875 SLGGRGVAVAGVGTAFAGGDAG--GMIGQGPMMAIRIAVASQSEEQEERAE 928

Oy 110 ---LP---RLPALYGECLTEAPRNRYRYQNVCISYSRWMDMARWE 151

Dd 929 SQSEEQEARAESFLPVVSARPVEVERADPRSSP-----EPLPWE 969

Oy 152 ----REI--DMWALNGIMLA----AWSQGEAIWORYYTL-----182

Dd 970 DIQGVSLVQRIDLSGDLEADDTISLISEVDPAVLNLSDIYDKYLPEFEEMIPIKVESKA 1029

Oy 183 ----GLTQAETNEF-----FGCP-----AFLAMGNMHTWDGP 213

Dd 1030 QEPSPSPFAEBELAEFPFTFWMPFGELGFPHGLEITESEBDVDLLLEAVGRKRKKSSP 1089

Oy 214 LPPSMHILKOQLQHRRVIDOKRSFGMTPLRALPAGHV-----PEAVTRVPQVNVTIKMG 266

Dd 1090 SRLFFHPGG--RHLPDEPAELGLRRVRKASVENHIRLKGRPEGKEKGPP---RKRP 1143

Oy 267 SWGHFNQSYSGSFLAAPDDPFIFITSLDELKLKEGTHIVADTFENMQPSSEPSY 326

Dd 1144 GLASFRLSGLSKWDRAP-----TFNEL-----SDETVV 1172

Oy 327 LAATTATAYEMTVTDTEAVNLLOGMWLFQHOPOFMGPBQAIRAVLGAVPRG--RLLVTDLF 384

Dd 1173 LGQSVTILACVYSAOPAAOATWSKDGAPLE-----SSSHVLIISAULKNFOLLTIIVVV 1224

Oy 385 AESQPVTNRITASFCGQPIFWCMILNPFGNHGLFGALEAVNNGGPAARLFRPSTWTGICMA 444

Dd 1225 AEDLCVYT---CSVINALGYTVTIGYLR---KAER--PESS-----1257

Oy 445 PEGISOHEEVVYSLMAELGWKRDVPVD-----LAAM-VTSFAARIYGVSHPDACA 493

Dd 1258 -----PCPDIGEYVADGVILLVMKREVESYGPTYIVQCSSLREGS 1295

Oy 494 WRLLRLSRVYNC-----SGEAC 509

Dd 1296 WTTLASDLFDCCYLTSKLSRGQTTPTRACYCKAKMGYPSSPSQOVYLLGSSHILASEBS 1355

QY 510 RGNHRSPLVRPRSLQNTMSIWNR--SDVFEARLLLTASPLATSPAFRYDLDTL--Q 566  
 DB 1356 QGRSAQPLPSTKTPAFQTOIGRFSVVRQCEKASGRLLAKIIPHPKDTAVLREYE 1415  
 QY 567 AVQELVSLYEERASAYLS--KELASLLPAGVLAAYELLPAIDVLAUSDSPFLGSMLEQA 625  
 DB 1416 ALKGLRHPLAQHLAAAYLSPRHVLILEL--CSGPPELLPCLAE--RASYSSESEVXDYLQM 1472  
 QY 626 RAAA-----VSEAEADFEYQN---SRVQLTLMGPEGNILDYANKQLAGLVANYTTPRRL 677  
 DB 1473 LSAIQLYLNQHTLHLDLRSEMMITTEYNL-----KVVDIGNAQSLSQEKVLPSEKFD 1526  
 QY 678 FLEALVDSVAQG---IPFOHQFDKNVFOI--EQAFVLSKORYPSPQPRG 721  
 DB 1527 YLETWAPPELLLEGQAVP-----QTDIWAIGVTAFLMSAEYVPSSEG 1568  
 RESULT 14  
 AAO15372 standard; Protein; 1665 AA.  
 AC AAO15372;  
 XX AAO15372;  
 DT 19-SEP-2002 (first entry)  
 XX 19-SEP-2002 (first entry)  
 DE Human myosin light chain kinase subfamily-related kinase protein.  
 XX Human myosin light chain kinase subfamily; kinase protein-mediated disease;  
 KW transgenic animal.  
 XX Homo sapiens.  
 CS WO200240683-A2.  
 XX PD 23-MAY-2002.  
 XX 22-OCT-2001; 2001WO-US32616.  
 XX 14-NOV-2000; 2000US-0711134.  
 PR 17-MAY-2001; 2001US-0858664.  
 XX (PEKE ) PE CORP NY.  
 PA Wei M, Ketchum K, Di Francesco V, Beasley EM;  
 PI WPI; 2002-500223/53.  
 DR N-PSDB; AAL43908, AAL43909.  
 PT New kinase proteins related to myosin light chain kinase subfamily and  
 encoding polynucleotide, useful for diagnosing, treating disease or  
 condition mediated by the kinase protein and for identifying modulators  
 PT  
 PS Claim 1; Fig 2; 96pp; English.  
 XX The invention comprises the amino acid and coding sequences (located on  
 CC chromosome 1) of a human kinase protein that is related to the myosin  
 CC light chain kinase subfamily. The human kinase DNA and protein sequences  
 CC of the invention are useful for identifying agents that modulate the  
 CC activity of the human kinase protein. Kinase-modulating agents are useful  
 CC for treating a disease or condition mediated by a human kinase protein.  
 CC The human kinase DNA sequences can be used to produce transgenic animals  
 CC which are useful for studying the function of kinase proteins and  
 CC identifying/evaluating modulators of kinase protein activity. The present  
 CC amino acid sequence represents the human kinase protein of the invention.  
 XX  
 SQ Sequence 1665 AA;  
 Query Match 3.0%; Score 118.5; DB 23; Length 1665;  
 Best Local Similarity 18.6%; Pred. No. 0.95;  
 Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

QY 19 GGAAGDEAREAAVAVLVARL--LGGPAAAD---FSVVERALAAKGLD-----TY 65  
 DB 815 GSAASQSSSQVSLVAGSSQVTEPSPIDNAGMTQAEHDLSDSPITLQRPQGYTKRF 874  
 QY 66 SLGG--GGAARVVRGSGTGAAGAALHRYLRDPCG-----CHVANGSQLR----- 109  
 DB 875 SLGGRGVAGVAGYGTFAFGDAG-----GMLGGQPMWARLMAVSSQSEEEQEARAE 928  
 QY 110 -----P---RPLPAVPGELTEATNRRYVQNVCTQSYSEVWMDARAE 151  
 DB 929 SQSEEQEARASPLPQVSARVPEVGRAPTRSSP-----EPTPE 969  
 QY 152 -----REI--DMALNGLNAL-----AMSGQALWQRYVAL----- 182  
 DB 970 DIGQVSLVQIRDLSDGAERADITSLDISVDPAVNLSDLYDKLPFEFMIFRVPKSA 1029  
 QY 183 -----GLTQALINF-----FTGP-----AFLANGRMGNLHTWGP 213  
 DB 1030 QEPSPMAHEELAEFPETFWPBGELGPHAGLEITSESEVDVALLAEANVGRKRWSSP 1089  
 QY 214 LPPSWHIKQLYLQHRVLDQMRSGMTPVLPAPAGHV-----PEAVTRVPQVAVTKWG 266  
 DB 1090 SRLTFHPG---RHLPDEPAELGRERVKASVEHISRLKGRPEGLEKEGPP---RKXP 1143  
 QY 267 SMGHENCSYSCSFLAPEDPIPTIGSLFLRLIKEFGTDHYGADTFEMQPPSSPEBY 326  
 DB 1144 GLASFRLSGKSMDBAP-----TFLRL-----SDEVTV 1172  
 QY 327 LAAATTAAVEAMTAVDTEAVWLLQGMFLFOHQPOFWGPAICRAVLGAVRG--RLVLDLF 384  
 DB 1173 LGQSVTLAQVSAQAQAQATWKGDAPL-----SSRVLISATLKNFGLTILVVV 1224  
 QY 385 AASQPYTFTASFOGQPTWCMLHNPNGNGLFGALLEAVNGSPEARLFPNSTMTGTGA 444  
 DB 1225 AEDLGYTT-----CSVSNAIGVTVTGYLR-----KAER--PSSS----- 1257  
 QY 445 PEGISQNEVYVLMELGWRKDPVPD-----LAAM--VTSFAARRYGSHPDAGA 493  
 DB 1258 -----PCPDIGEVADSVLWTKPVESTGYPTIYQCSLEGS 1295  
 QY 494 WRLLRSVYNC-----SGEXC 509  
 DB 1296 WTLTASDIPDCCYLTSKLSRGQTYTFRVACVSKAGGFPSPSEQVLLGSPHLSSEBS 1355  
 QY 510 RGNHRSPLVRPRSLQNTMSIWNR--SDVFEARLLLTASPLATSPAFRYDLDTL--Q 566  
 DB 1356 QGRSAQPLPSTKTPAFQTOIGRFSVVRQCEKASGRLLAKIIPHPKDTAVLREYE 1415  
 QY 567 AVQELVSLYEERASAYLS--KELASLLPAGVLAAYELLPAIDVLAUSDSPFLGSMLEQA 625  
 DB 1416 ALKGLRHPLAQHLAAAYLSPRHVLILEL--CSGPPELLPCLAE--RASYSSESEVXDYLQM 1472  
 QY 626 RAAA-----VSEAEADFEYQN---SRVQLTLMGPEGNILDYANKQLAGLVANYTTPRRL 677  
 DB 1473 LSAIQLYLNQHTLHLDLRSEMMITTEYNL-----KVVDIGNAQSLSQEKVLPSEKFD 1526  
 QY 678 FLEALVDSVAQG---IPFOHQFDKNVFOI--EQAFVLSKORYPSPQPRG 721  
 DB 1527 YLETWAPPELLLEGQAVP-----QTDIWAIGVTAFLMSAEYVPSSEG 1568  
 RESULT 15  
 AAB30569 standard; Protein; 2596 AA.  
 AC AAB30569;  
 XX AAB30569;  
 DT 19-MAR-2001 (first entry)  
 XX A splice variant of a signal transduction polypeptide.  
 DE Signal transduction; H19G5; kinase; cardiac disease; angina pectoris;  
 KW

KM congestive heart failure; dilated congestive cardiomyopathy;  
 KM hypertrophic cardiomyopathy; restrictive cardiomyopathy; hypertension;  
 KM mitral valve disease; aortic valve disease; tricuspid valve disease;  
 KM myocardial infarction; cardiac arrhythmia; arteriosclerosis;  
 KM atherosclerosis; cardiac tumour; microbial infection; splice variant.  
 XX Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT Domain 325..504  
 FT /note="guanine nucleotide exchange factor domain"  
 FT Domain 1094..1351  
 FT /note="kinase domain"  
 FT Domain 2301..2553  
 FT /note="kinase domain"  
 XX  
 XX MO200063381-A1.  
 XX  
 XX PD 26-OCT-2000.  
 XX  
 XX PF 11-APR-2000; 2000MO-US09488.  
 XX  
 XX PR 16-APR-1999; 99US-0129553.  
 XX  
 XX PA (SCIO-) SCIOS INC.  
 XX  
 XX PI Zeng W, Stanton L, Kong H;  
 XX  
 XX DR WPI; 2001-007013/01.  
 XX  
 XX DR N-PSDB; AAC62287.  
 XX  
 PT Novel h19G5 polypeptides capable of regulating signal transduction and  
 PT exhibiting kinase activity useful for identifying antibodies to treat  
 PT cardiac diseases, and additional mediators of signal transduction -  
 XX  
 PS Claim 1; Page 68-74; 81pp; English.  
 XX  
 CC The present sequence represents a splice variant of human in signal  
 CC transduction polypeptide. The polypeptide is designated H19G5. The  
 CC protein is capable of regulating signal transduction and exhibits kinase  
 CC activity. The H19G5 transcript is expressed in the heart. H19G5  
 CC polypeptides and polynucleotides are useful for preventing or treating a  
 CC cardiac disease, such as congestive heart failure, dilated congestive  
 CC cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy,  
 CC mitral valve disease, aortic valve disease or tricuspid valve disease,  
 CC angina pectoris, myocardial infarction, cardiac arrhythmia, pulmonary,  
 CC arterial or rheovascular hypertension, arteriosclerosis, atherosclerosis  
 CC and cardiac tumours in humans. The polypeptide is also useful for  
 CC detecting the expression of a protein capable of regulating signal  
 CC transduction or the expression of a protein capable of acting as a donor  
 CC or acceptor molecule of a phosphate group. The monoclonal antibodies can  
 CC be used as probes for detecting discrete antigens expressed by tissue or  
 CC cell samples, and therefore used in humans for localization and  
 CC monitoring of microbial infection.  
 CC  
 XX  
 SQ Sequence 2596 AA;  
 Query Match 3.0%; Score 118.5; DB 22; Length 2596;  
 Best Local Similarity 18.6%; Pred. No. 1.9; Indels 319; Gaps 40;  
 Matches 165; Conservative 97; Mismatches 307;  
 QY 19 GGAAGDEARBAAVRALVARL-LGPGPAD--FSVSEVERALAKPGLD-----TY 65  
 DB 1746 GSASQASSSQVSSLRVSSQVGTPEPSPILDAEGMTQEAEDLSSTPTLQRPQOATMRKF 1805  
 QY 66 SLGG-GGAARVRVNGSGVAAAAGLHRYLRDPG--CHVAMSGSLR----- 109  
 DB 1806 SLGGRGVAGVAGTATFAGDAG-----GMLGGPMARIKAVSQQSEEEQEARAF 1859  
 QY 110 -----LP-----RPLPAVPGELTEATPNRRYRYQNVCTQSYSPFWMDMARWE 151  
 DB 1860 SQSEEQEQARAEAPLPQVARSAPVEVGRAPTRSSP-----EPTPWE 1900

QY 152 -----REI--DWMALNGINLAL-----AMSGQEAIMQREVYAL----- 162  
 DB 1901 DIQVSLVQIRDLSDGDAEADDTSLDISEVDPAVNLSDLYDKYLPFEEMIPRKYPKSA 1960  
 QY 183 -----GLTQAEINEF-----FTGP-----AFLAMGMLHMTWDGP 213  
 DB 1961 QPEPPSPMAEEBELAEFEPEPTWMPGELGPHAGLEITEESBDVLLAEAAVGRKKKSSP 2020  
 QY 214 LPPSMHIKQLYLQHRVLDQMSFGMPEVLPAPAGHV-----PEAVTRVPQVNTXMG 266  
 DB 2021 SRSLFHPG---RHLPLDEPAELGLRRVKAASVEHSRIILKGRPEGEKSGP---RKKP 2074  
 QY 267 SWGHFNCSYSCSFLLAEDPIPIIGSLFRELKERGTHTHYGADTFNMQSPSSPSY 326  
 DB 2075 GLASFRLSGKSWDRAP-----TFPLREL-----SDETVV 2103  
 QY 327 LAAATTAVEAMTAVDTEAVWMLQGMWLFQHPQPFMGPAQIRAVLGAVPBG--RLVLDLF 384  
 DB 2104 LGQSVTLACQVSAQPAQATMSKDGAFL-----SSSRVILSNATLKNFOLLTLVV 2155  
 QY 385 AESQPVYTRTASFGQOPFIWCMLEHNFQGNHGLFGALBAVNGGPBARLFPNSTMTVGMA 444  
 DB 2156 AEDLGVT-----CSVSNALGVTITGVLR-----KAER--PSSS----- 2188  
 QY 445 PEGISQNEVYYSIMAEIGWKKDPVPD-----LAAM--VTSPARRYGVSHPDAGAA 493  
 DB 2189 -----POPDIGEYADGVLLVMKPVESYGPVYTYIQCSLEGSS 2226  
 QY 494 WRLLRSVYNG-----SGEAC 509  
 DB 2227 WTLASDIFQCCVYLSKLSRGTYFRTACVSKAGMPYSPSEQVLLGAPSHLASEES 2286  
 QY 510 RGHRSPLVRRPSIQMNTSIWNR--SDVFEAMRLLTSASLATSAPFRYDLDLTR--Q 566  
 DB 2287 QGRSAQPLPSTKTAFTQTQIRGRFSVVRQCMERASGRALAAKIIYPHPADKRAVLYREY 2346  
 QY 567 AVCELVSLVVEEARSAVLS--KEIASLRAGGVLAAYELLPALDEVLASDSRFLGSMLEQA 625  
 DB 2347 ALKGLRHPHLAQHAATLSRHLVLIEL--CSGPELLPCLAE--RASYSSEVVDYLMQM 2403  
 QY 626 RAAH-----VSEADADYEON--SRQYLMPGEGNILDYANKQLAGVANYTTPWRLL 677  
 DB 2404 LSAQYILNHQIHLILDRSENMITEYNLL-----KVVDLGNAGSISQEKVLPSSDKFKD 2457  
 QY 678 FLEALVDSVAGQ---IPFOQHOPDKNVFOL--EQAFVLSKORYSQPRG 721  
 DB 2458 YLEIMARPELLBGQAVP-----QTDIWAIGVATFAFIMLSAEPVSSBG 2499

Search completed: February 13, 2004, 16:15:23  
 Job time : 60 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2004, 15:44:39 ; Search time 29 Seconds  
(without alignments)

2463.908 Million cell updates/sec

Title: US-09-836-613-2

Perfect score: 3939

Sequence: 1 MEAVAVAAAVGVLLAGAGS.....VDLAKKIFLKYYRGMVAGSW 743

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR\_76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3939	100.0	743	1 G02270	alpha-N-acetylgluc
2	1060.5	26.9	770	2 G87316	alpha-N-acetylgluc
3	123.5	3.1	497	1 G02938	probable debrisoqu
4	121	3.0	4342	2 H83343	probable non-ribos
5	119	3.0	3649	1 S18268	delta-(L-alpha-aml
6	117.5	3.0	497	1 Q4HED1	debrisoquine 4-hyd
7	117	3.0	1374	2 D72593	hypothetical prote
8	115.5	2.9	973	2 T35238	probable secreted
9	115	2.9	725	2 C87485	ComEC/Rec2 family
10	114	2.9	479	2 T31293	aldelyde dehydroge
11	113	2.9	1281	2 G83405	hypothetical prote
12	112	2.8	1323	2 I78557	N-methyl-D-asparta
13	111	2.8	1027	2 H87316	hypothetical prote
14	109.5	2.8	285	2 F87285	conserved hypotet
15	109.5	2.8	814	2 T47641	hypothetical prote
16	109.5	2.8	1377	2 G55159	thsa protein precu
17	109.5	2.8	1411	2 B55145	thsb protein precu
18	109	2.8	1080	2 T19048	probable pro-x car
19	107.5	2.7	1622	2 D86428	glutathione S-conj
20	107	2.7	741	2 A83271	hypothetical prote
21	106.5	2.7	666	2 A87577	oligopeptid trans
22	106.5	2.7	1678	2 T35547	hypothetical prote
23	104.5	2.7	726	1 G5ECP	catalase (EC 1.11.
24	104.5	2.7	726	2 G81237	hydroperoxidase HP
25	104	2.6	1230	2 T07863	soluble starch syn
26	103.5	2.6	726	2 C86085	catalase, hydroper
27	103.5	2.6	1737	2 T00209	MEGF8 protein - hu
28	103	2.6	698	2 D90771	hypothetical prote
29	103	2.6	698	2 H85633	hypothetical prote

30	103	2.6	1400	2 B70963	hypothetical prote
31	102.5	2.6	1623	2 T01369	ABC transporter At
32	102	2.6	352	2 B87507	conserved hypotet
33	102	2.6	435	2 T35884	tagatose 6-phospha
34	102	2.6	9376	2 T14593	syringomycin synth
35	101.5	2.6	638	2 AD3111	myo-inositol catab
36	101.5	2.6	638	2 H98175	myo-inositol catab
37	101.5	2.6	843	2 F96537	hypothetical prote
38	101.5	2.6	3194	2 D71517	toxoin-like outer m
39	101.5	2.6	5069	2 T17464	rifamycin polyketi
40	101	2.6	698	2 F64839	yea protein precu
41	101	2.6	709	2 B82580	alanyl dipeptidyl
42	101	2.6	3413	2 T17467	rifamycin polyketi
43	101	2.6	3519	2 S43048	polyketide synthas
44	100.5	2.6	4735	2 T17463	rifamycin polyketi
45	100	2.5	726	2 A10936	catalase (hydroper

## ALIGNMENTS

## RESULT 1

G02270 alpha-N-acetylglucosaminidase (EC 3.2.1.50) - human

C/Species: Homo sapiens (man)

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #ext\_change 03-Dec-1999

C/Accession: G02270

R/Weber, B.; Scott, H.; Hopwood, J.J.

submitted to the EMBL Data Library, November 1995

A/Reference number: H00952

A/Accession: G02270

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-743 <WEB>

A/Cross-references: EMBL:U40846; NID:g1197839; PID:g1197840

A/Genes: GDB:NAGLU

A/Cross-references: GDB:636533

A/Map position: 17c21

A/Note: deficient in Sanfilippo B syndrome

C/Superfamily: human alpha-N-acetylglucosaminidase

C/Keywords: glycosidase; hydrolase

Query Match	100.0%	Score 3939	DB 1	Length 743
Best Local Similarity	100.0%	Pred. No. 5e-277		
Matches 743	Conservative	0	Mismatches 0	Indels 0
				Gaps 0
QY	1 MEAVAVAAAVGVLLAGAGAGDARBAANRALVARIIGPRAADTSVVERLAAP	60		
DB	1 MEAVAVAAAVGVLLAGAGAGDARBAANRALVARIIGPRAADTSVVERLAAP	60		
QY	61 GLDTYSIGGGGAARVRVSGSTGVAAAAGHRLRFQCGCHVAMSSQRLPPLPAVPG	120		
DB	61 GLDTYSIGGGGAARVRVSGSTGVAAAAGHRLRFQCGCHVAMSSQRLPPLPAVPG	120		
QY	121 LREATPNRRYYQNYCTOSYSFVWMDMARWEIEDMALINGINALAMSGDAINQRYL	180		
DB	121 LREATPNRRYYQNYCTOSYSFVWMDMARWEIEDMALINGINALAMSGDAINQRYL	180		
QY	181 ALGLTQAEINERFPTGPAFLAMGRMGLHTWQDPLPSSHIIKQLYCHVLDOMRSGMT	240		
DB	181 ALGLTQAEINERFPTGPAFLAMGRMGLHTWQDPLPSSHIIKQLYCHVLDOMRSGMT	240		
QY	241 VLPAPAGHVEAVTRVFPQVNTKXSGHFNCSYSCSFLAPEDPIFPIIGSLFRL	300		
DB	241 VLPAPAGHVEAVTRVFPQVNTKXSGHFNCSYSCSFLAPEDPIFPIIGSLFRL	300		
QY	301 KEFGTDHITGADTFNEMQPSSEPSYLAATTAAYEANTAVDTEVWLQGLPFOHQ	360		
DB	301 KEFGTDHITGADTFNEMQPSSEPSYLAATTAAYEANTAVDTEVWLQGLPFOHQ	360		
QY	361 WQPAQIRAVLGAIVPGRLLVLDLFAESQPVYTRTASFOGPIFWMGLNFGNHLFGAL	420		
DB	361 WQPAQIRAVLGAIVPGRLLVLDLFAESQPVYTRTASFOGPIFWMGLNFGNHLFGAL	420		

Db 361 WSGAQRRAVLGAVPRGRLLVLDLFAESQPYTRTASFGQPFIMWMLHNGHGLFGAL 420

Qy 421 EAVNGPEEARLPFNSTMTGTGMAPRGISQNEVYVSLMAELGRKDPVVDLAAMTTSFA 480

Db 421 EAVNGPEEARLPFNSTMTGTGMAPRGISQNEVYVSLMAELGRKDPVVDLAAMTTSFA 480

Qy 481 RRVGVSHPDGAAMRLLRGVNCSGACGHNRSPLVRRPSLQMTSITWYRSDVFEAM 540

Db 481 RRVGVSHPDGAAMRLLRGVNCSGACGHNRSPLVRRPSLQMTSITWYRSDVFEAM 540

Qy 541 RLILTSAPSLATSPPARYDLDLITRQAVQELVSLYEARSAYLSKELASLLRAGCVLAY 600

Db 541 RLILTSAPSLATSPPARYDLDLITRQAVQELVSLYEARSAYLSKELASLLRAGCVLAY 600

Qy 601 ELIPALDEVLASRSRFLGSMLEQARAASVSEADPFYQNSRYOULTMGPEGNILDYAN 660

Db 601 ELIPALDEVLASRSRFLGSMLEQARAASVSEADPFYQNSRYOULTMGPEGNILDYAN 660

Qy 661 KOLAGLVANYTTPRMRFLFELALVDVAQGIFFQOHQFDKNVFLQEQAFVLSKQRYPSQPR 720

Db 661 KOLAGLVANYTTPRMRFLFELALVDVAQGIFFQOHQFDKNVFLQEQAFVLSKQRYPSQPR 720

Qy 721 GDTVDLAKKIFLKYPRGVWAGSW 743

Db 721 GDTVDLAKKIFLKYPRGVWAGSW 743

## RESULT 2

alpha-N-acetylglucosaminidase [imported] - Caulobacter crescentus  
C87316  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: C87316  
B:Nieman, W.C.; Fealduym, T.V.; Paulsen, I.T.; Nelson, K.B.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gilm, M.L.; Haft, D.H.; Kojot  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: C87316  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-770 <SRO>  
A:Cross-references: GB:AE005673; NID:g13421729; PIDN:AAK2527.1; GSPDB:GN00148  
C:Gene: CC0540  
C:Superfamily: human alpha-N-acetylglucosaminidase

Query Match 26.9%; Score 1060.5; DB 2; Length 770;  
Best Local Similarity 33.0%; Pred. No. 1,2e-68;  
Matches 260; Conservative 132; Mismatches 302; Indels 93; Gaps 19;

Qy 2 EAVVAVAAVGVLLIAGAGAGDBREAAVRLVARLLGPGPADFSVSEBRLAAKGG 61

Db 11 QALSLAAAVASPALAAGSTG---VAAARASLKLKFLGRRLAG-----AHLVTG 59

Qy 62 LDT--YSLGGGAGARVVRSTGVAAAGLHRYLRDFCGGVAWSGSLRLPRLPAPVG 119

Db 60 AENSWVIAIGKGA--IISGDSPALVRGAYAHLRQAGLAHVWEDRVQAGAVRAGG 118

Qy 120 ELTEAPNRYRYQNVCTQSYSFVWMDAWEERIDMMLANGINLAWSGQALWQRY 179

Db 119 ARVE--TFPRHAYLNTCTGYTTPWGMGWRTEIMMAAHGIDMLAMEGGEYVRRALM 177

Qy 180 LAIQLTQAEINEPFTGAPFLAMGNGMLHTWDGLPLPSMHIKQLYLOHVLIDMRSGMT 239

Db 178 REFSLAEALADIFSGAFTPMHMGNIQYKAPLPIAMDKKDLQVKILGMRSLGNT 237

Qy 240 PVLPAFAGVPEAVTRVFPQVNTYKSGWGFNCYSQSFLLAPEDPIPIIGSLFLREL 299

Db 238 PIIIPAFGVYFKAAEKNPARIYRMPEGFHEFY---WIDPADPLFAKIAARFLALY 293

Qy 300 IKRGTHIYIGADITNEMQPP-----SSEPSY-----LA 328

Db 294 TETFGAGTYIADSFNEMLPINADGADDAVQSTANTATYTKKVEVDPALKAQRLA 353

Qy 329 AATAVTEAMTAVDTAENVLLIQMLFQHPQFMGPAQIRAVLGAVPRGRLLVLDLFAESQ 388

Db 354 AYGAIVDSIRKOTRPPDAVVMWQGLFGADSHFNDPAISAYSLVLPDDKLMILIDIGDRY 413

Qy 389 P-VYTRTASFGQPFIMWMLHNGHGLFGAL-----EAVNGPEEARLPFNSTMTVG 440

Db 414 PNWKAKAFQGGKMTIYGVYHANGSNIPYVGDGFGFYQDIPALANPDACKL-----AG 467

Qy 441 TGMAREGISQNEVYVSLMAELGRKDPVVDLAAMTTSFAARRGVSHPDGAAMRLIRS 500

Db 468 FGMPEEGIHNNISIYEAIVYDLAMSEQASP-ATWLTYYAARVYGKTSFALDALGQLVFA 526

Qy 501 VYNGSGEACGHNRSF---LVRRPSLQMTSITWY--NRSDVFEAMLLTSAPSLATSP 554

Db 527 AFSTRWSPRWKSKAQAAYLFKKPIATVGDFFQHPGDRAKLEAAVYALTLALPTYGQPR 586

Qy 555 AFRYDLDLITRQAVQELVSLYEARSAYLSKELASLLRAGVLAZLLPALDEVLASDS 614

Db 587 LFVLDLTDATRHLATMKIDLLQVAVAAYRRGDTA---AGDAARVSI-----EALALSI 637

Qy 615 RFLG-----SWLEQARAASVSEADPFYQNSRYOULTMGPEGNILDYANKOLAGLV 667

Db 638 DKLGVQPDLTATVIDERAFAGDTPADAAIVAAKAKQVITWGGGNLNDYASAKWGLY 697

Qy 668 ANYTTPRMRFLFELALVDVAQGIFFQOHQFDKNVFLQEQAFVLSKQRY-----PSQPRGDT 723

Db 698 KSFLPFRMSRLDL--KAAGTGFDEVYTRGVAMERAMVEEVAVREKPADPIGEI 755

Qy 724 VDLAKKI 730

Db 756 KTLIARI 762

## RESULT 3

G02938  
Probable debrisoquine 4-hydroxylase (EC 1.14.14.-) cytochrome P450 - crab-eating macaque  
C:Species: Macaca fascicularis (crab-eating macaque)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Mar-2000  
C:Accession: G02938  
B:Lawton, M.P.; Laddison, K.J.; Speirs, A.A.; Mankowski, D.C.; Tweedie, D.J.  
submitted to the EMBL Data Library, October 1995  
A:Reference number: G12616  
A:Accession: G02938  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-497 <LAW>  
A:Cross-references: EMBL:U08218; NID:g1022899; PIDN:AAA79722.1; PID:g1022900  
C:Genetics:

Query Match 3.1%; Score 123.5; DB 1; Length 497;  
Best Local Similarity 22.9%; Pred. No. 0.45;  
Matches 104; Conservative 47; Mismatches 140; Indels 163; Gaps 22;

Qy 343 TEAVWLL-----QGLFQHPQFMGPAQIRAVLGAVPRGRLLVLDLFAESQPYTRT 394

Db 12 TVALFLLVLDLMHRRQMAARYP--GPIPLPGL-----GNLHVD----- 50

Qy 395 ASFGQPRIMWMLHNGFN-----HGLFGALFA-VNGPEEARL--FNST 437

Db 51 --FKNTFYCFQLRRRGVNSVSLQNTAPVYVNLGLAAVREALYTCEDTRDRPPVYINQ 108

Qy 438 MVGTGMAPEGI-----SQNEVYVSLMAELGRKDPVVDLAAMTTSFAARRGVSH 487

Db 109 VLGQPRSGQGVFLARYGPAWREGRFSVSTLRNLGLK---KSLQWVTEBAACLCAAFT 165

QY 488 PDAGAMR--LLRSVYNGSGEACGRHNRSPLVRPSLQNTS:TWYNSDVFEAMRLLL 544  
 DB 166 DQAGRPFRPNSLDDKAVSN-----VVASLTYGR----- 193  
 QY 545 TSAPLAPSPAPRYD-----LLDLTRQAVQELVSLYEEARSANLSELAS---LLR-- 593  
 DB 194 -----REYDDPRLFLFDLTHLELKE-----BSGFLREVLNAIPLLLRIP 234  
 QY 594 --AGVLAWE--LLPALDEVL-----ASDRSFLGLSWLEQARAAVAEADPYEQ 640  
 DB 235 GLAGVLSQKAFLLQDLBELTEHRTWDPQAPRDLTEAFLEMEKAK--GNPSSFNEE 293  
 QY 641 NSRYQLTMGPENGLIDPAKQLAGLVANYTTPRWRLEALVDSVACGIFQGHQDPKN 700  
 DB 294 NLR-----MVVADLFSAQMVTSTTLAAGLLMLTHPDVQRRV---QOEIDDV 338  
 QY 701 VFQLEAFVLSKORYP-----SQPRGDTVDL 726  
 DB 339 IGQVRPBMGDQARMPYTAVTHVEYRGDVIPL 372

## RESULT 4

H83343  
 C:Probable non-ribosomal peptide synthetase PA2424 [imported] - Pseudomonas aeruginosa (sc  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: H83343  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; Llm,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: A82950; PMID:20437337; PMID:10984043  
 A:Accession: H83343  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-4342 <STO>  
 A:Cross-references: GB:AE004669; GB:AE004091; NID:g9948460; PIDN:AA05812.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA2424  
 C:Superfamily: acyl carrier protein homology; acetate-CoA ligase homology  
 C:Keywords: carrier protein; phosphopantetheine; phosphoprotein  
 F:60-553/Domain: acetate-CoA ligase homology <ACLI>  
 F:584-652/Domain: acyl carrier protein homology <ACP1>  
 F:1174-1622/Domain: acetate-CoA ligase homology <ACI2>  
 F:1637-1705/Domain: acyl carrier protein homology <ACP2>  
 F:2232-2689/Domain: acetate-CoA ligase homology <ACI3>  
 F:2708-2773/Domain: acyl carrier protein homology <ACP3>  
 F:3759-4230/Domain: acetate-CoA ligase homology <ACI4>  
 F:4248-4316/Domain: acyl carrier protein homology <ACP4>  
 F:1669,2738,4280/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 3.1%; Score 121; DB 2; Length 4342;  
 Best Local Similarity 21.4%; Pred. No. 14;  
 Matches 178; Conservative 101; Mismatches 318; Indels 236; Gaps 44;  
 QY 4 VAVAA-----AVGTLTAGAGCA-----AGDEARAAVRA-LVARLL 40  
 DB 1181 VAICERSPOLLVGLALIVKAGAVPLDPYPSERLALMLADSGVELLTLQALPRL- 1239  
 QY 41 GPG-----PAADFVSVERALAAKRGDLTYSLGGGGAARVVRSGTGYAAAAGLHRLRD 95  
 DB 1240 -FGAGVTPTICLDSIKLNNWPSQAFGLH--GNLAVITTSQSTQPKVG----- 1289  
 QY 96 FCGCHVANSQQLRPLRPAPVPG--LTEATPNRYRYQVNCQSYSFVWMDWARERE 153  
 DB 1290 --NTHAALA-ERLQWQATYTLIDGDDVLMQAKP-----VSFVSVW--E 1328  
 QY 154 IDMMALANGINIALAMSG-----EAIQPVIALGLTQAEINFEFTGAPLAWRMG 205  
 DB 1329 CFWPLVTCRLVLAAPGHRDPARLVELVROFGVTTLHFVPLQLFIDEGVAACSLR 1388

QY 206 NHTWQDPLPSPWHIQLYQHRVLDQMSFGMTPLYPAFAGVPAVTRVFPQVNTKX 265  
 DB 1389 RLFSGEALPAE-----LNRVLQR-----LPVALHNRYGPTET--AINVT-- 1428  
 QY 266 GSWGHNCSSCSFLAPD-----PIPIIGSLFLBELKEPETHITYGADTFNEN----- 317  
 DB 1429 ---HWQC-----RAEDGERSPIGRPLGNVCVLAERF--NLLPFGVAGELCIG 1473  
 QY 318 -----QPSSEPSYLAATTAVYEAMTAVTEAVMLQGM-----FOHOPWG- 362  
 DB 1474 LGLARVGLRPAISARFVADPESAAGERLYRTGDRANNAQVLEYLRLDQVYLRGF 1533  
 QY 363 ---PAQIRAVLAVP---RGLLVLDLPABSDPY--YTTTASQOQPFITCMHNGNA 414  
 DB 1534 RIEPBEIQARLLAQPVAAVAVVIRGVAGSOLVGYYTAVAGAEAA-----EQHQ 1584  
 QY 415 GLFGALAVNGGPEARLPFNSTMTVGTGAPEGISONEVYYSIMAEIGMRK-----DPVPD 470  
 DB 1585 RLRAALQA-----ELDEVVPTQLMRLAQMPGPS--GKLDTRALPEPVQOQREHVEPTE 1638  
 QY 471 L-----AAVTSFAARRYG-----VSHPDGAAMRLLSVYNGS-- 505  
 DB 1639 LQRIIAIWSBYLGLPRVGLRDPFELGHSILATRIVSRTQACDVELPALFEASEL 1698  
 QY 506 -----GEACGRHNRSPLVRPSLQNTS:TWYNSDVFEAMRLLLSAPSATSAPF 556  
 DB 1699 EAFCEQVRAQAAGRDSGAIIRIRIQEPVPVLSQGMFLWL-----EPD--SPAY 1751  
 QY 557 KY-----DLADLTR--QAVQELVSLYEEARSAYLSKELASLRA---GGV-LAYELL 603  
 DB 1752 NVGGLARLSPDVAFAFALQVQ--RHETRTTPPSVDGVQVHGDGILHMDWQDF 1810  
 QY 604 PLAD-----EVLASDRFLIGSLRQA---RAAASEADPFQMSRQQLTMGPE 652  
 DB 1811 SALDRDSRQOHLQTLA-DEBARHPDLESGPLIRVCVMAREERHLVTLHHIV--TE 1866  
 QY 653 GNILDYANKQLAGLVANYTTPRWRLEALVDSVACGIFQGHQDPKNVFOLE 705  
 DB 1867 GMMMDIFAREIGLYEAFIDDR-----ESPLEPLVQ--YLDYSWQRE 1908

## RESULT 5

S18268  
 delta-(L-alpha-aminoadipyl)-L-cysteiny-L-D-valine synthetase - Streptomyces lactamdurans  
 C:Species: Streptomyces lactamdurans  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Nov-2000  
 C:Accession: S18268; B38171  
 R:Martin, J.F.  
 submitted to the EMBL Data Library, January 1991  
 A:Reference number: S18268  
 A:Accession: S18268  
 A:Molecule type: DNA  
 A:Residues: 1-3649 <MAR>  
 A:Cross-references: EMBL:X57310; NID:g45005; PIDN:CAA40561.1; PID:g45006  
 R:Coque, J.U.R.; Martin, J.F.; Galzada, J.G.; Liras, P.  
 Mol. Microbiol. 5, 1125-1133, 1991  
 A:Title: The cephamycin biosynthetic genes pcbAB, encoding a large multidomain peptide s  
 genes in Acetomycetum chrysogenum and Penicillium chrysogenum.  
 A:Reference number: S15283; PMID:92065808; PMID:1956290  
 A:Accession: S15283  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 224-940;1319-2010;2373-3307 <COQ>  
 A:Cross-references: EMBL:X57310  
 A:Note: the source is designated as Nocardia lactamdurans  
 R:Coque, J.U.R.; Liras, P.; Laiz, L.; Martin, J.F.  
 J. Bacteriol. 173, 6258-6264, 1991  
 A:Title: A gene encoding lysine 6-aminotransferase, which forms the beta-lactam precurs  
 A:Reference number: A38171; PMID:92011390; PMID:1917857  
 A:Accession: B38171  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-23 <CO2>

A:Cross-references: GB:S57006  
 C:Genetics:  
 A:Gene: pcbaB  
 C:Superfamily: alpha-aminoacyl-tRNA synthetase; acetyl-CoA ligase homolog  
 C:Keyword: carrier protein; cephamycin biosynthesis; phosphopantetheine; phosphoprotein  
 F:786-758/Domain: acetate-CoA ligase homology <ACLI>  
 F:786-856/Domain: acyl carrier protein homology <ACPI>  
 F:1392-1844/Domain: acetate-CoA ligase homology <ACLI>  
 F:1862-1932/Domain: acyl carrier protein homology <ACPI>  
 F:2446-2895/Domain: acetate-CoA ligase homology <ACLI>  
 F:2942-2980/Domain: acyl carrier protein homology <ACPI>  
 F:520,1896,2944/Binding site: phosphopantetheine (Set) (covalent) #status predicted

Query Match 3.0%; Score 119; DB 1; Length 3649;  
 Best Local Similarity 21.1%; Pred. No. 16;  
 Matches 174; Conservative 63; Mismatches 283; Indels 306; Gaps 37;

16 AGAGGAAGDARAAAVRALV-----ATLGGPAPADSVSEVALAKPGLDYSIGG 69  
 1988 AALPGALGLVRRHPLRLTKTDGVRROYFIPADVRLVLP----- 2031  
 70 GGAARVRGSGTVAAGLHRLDFCGCHVAMSGQLRPPPLPVGSLTEATPMRY 129  
 2032 -----STVDRALDELVLT-----RAGYVRLHELP-----RAEFPHGD 2070  
 130 RYQNVCTGSGYVWMDARMERET-----DMALGINIALAMSGOEAIWQRY 179  
 2071 EYLSVYVHSCPDGSMWDFRRELALLDGVPEADLGALRGT-----YGEFAWQRY 2124  
 180 LALGLQAEINEFPTAPFLAMGKMGHMT-----MDGFLPSPMHK-----OLYGHRYLD 231  
 2125 LT-GKLLALTEWT-----GALGGFTIALPDHPFPFDRGRHLELDELRTTE 2176  
 232 QMSFGMT----- 239  
 2177 ALRELARTARVSLVYLGLAMCLMNTYGOHDTGVTSPARGRPEEPRAVGFANLLA 2236  
 240 -----PYLPAGHVEAVTRVPOVNVTKMSGWGFNCSYCSFLAEDPI-FPI 290  
 2237 LKRVDPAAITLPAYVSGEAV--VAAQV-----GELPREOLVKELKEDSRHPI 2287  
 291 IGLFLRELKKEFGDTHIGADTFENMOPSSSEPSYLAATTAAYEAMTAVDEAVWLQ 350  
 2288 LGNFTLQNV-----SDH-----TSALTGYQDS-----GGMTTKED-LSATWTERATGLA 2333  
 351 GML-----FOHQPFQPAQIRAVLAVGRGLVLDLFAESQ-FYV 391  
 2334 GNLTYAASFDDTSASGFIATFQHV-----LAEFSAQAQTPAQITALDEPQALPDA 2388  
 392 TRTASFQGFITWCMILHFGNGLFGALEAVNGCFEARLFPNSTWVGTG---MAPGI 448  
 2389 TRRARPPGP-----GRCTRLE-----EVAATMPDVAVAVHGVRLTYREL 2430  
 449 SQ-NEVYSLMAELMKRDP-----VPLAAMTSPARRRYGVSHPDAGAA 493  
 2431 NESANLHRLKRVAPRDELIALVLDKSELTVAIILAVWGAAYMIDPSYDDRIA 2490  
 494 W-----RLLRSVNCGEACRGNRSPVLRPELQNTSITWNRSPVFAMLLTS 546  
 2491 FMLSDFGALV-----AGEAHGSRVRC-----LTSGVDLDELQDLTG 2529  
 547 APSLATSPARRVLDLDTROAVGELVSLYEARSAYLSKELASLIRAGGLVAYELLPL 606  
 2530 EP--AANP-----VTETSTELAYAYI-----TSQTKPKPRAVIVSHGSV----- 2567  
 607 DELVADSDFLGSWLEQARAASVSEADFEQNRRYQTLTGPEGNTL----- 656  
 2568 DSRRAQSGRYFSGSPESAVALFL--ANYVPDEVEGLALSLVGHGLVLPSPSADD 2624  
 657 ---DYANKQAGLVANYTTPRWRLFLALVDSVAGGIFQHQFD 698  
 2625 PAVELANRE--GL-----SYLSGTPTQVERFD 2650

RESULT 6  
 C4HND1  
 debrisoquine 4-hydroxylase (EC 1.14.14.1) cytochrome P450 2D6 - human  
 N:Alternate names: CYP2D6; cytochrome P450 isozyme 2D; cytochrome P450db1  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Mar-1992 #sequence, revision 31-Mar-1992 #text, change 03-Mar-2000  
 C:Accession: S01199; A28883; J04156; A33629; A30335  
 R:Gonzalez, F.J.; Skoda, R.C.; Kimura, S.; Umeno, M.; Zanger, U.M.; Nebert, D.W.; Gelbo  
 Nature 331, 442-446, 1988  
 A:Title: Characterization of the common genetic defect in humans deficient in debrisoqu  
 A:Reference number: S01199; PMID:8122614; PMID:3123997  
 A:Accession: S01199  
 A:Status: translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-497 <CON>  
 A:Cross-references: EMBL:X08006; NID:G30450; PIDN:CAA30807.1; PID:G30451  
 R:Gonzalez, F.J.; Valbois, F.; Hardwick, J.P.; McBride, O.W.; Nebert, D.W.; Gelboin, H.  
 Genomics 2, 174-179, 1988  
 A:Title: Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino acid sequ  
 A:Reference number: A28883; PMID:88314109; PMID:3410476  
 A:Accession: A28883  
 A:Molecule type: mRNA  
 A:Residues: 1-497 <CON2>  
 A:Cross-references: EMBL:M20403; NID:G181349; PIDN:AAA52153.1; PID:G181350  
 R:Jiang, Q.; Voigt, J.M.; Colby, H.D.  
 Biochem. Biophys. Res. Commun. 209, 1149-1156, 1995  
 A:Title: Molecular cloning and sequencing of a guinea pig cytochrome P4502D (CYP2D16) :  
 A:Reference number: J04153; PMID:95251703; PMID:7733969  
 A:Accession: J04156  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-497 <RNA>  
 R:Kimura, S.; Umeno, M.; Skoda, R.C.; Meyer, U.A.; Gonzalez, F.J.  
 Am. J. Hum. Genet. 45, 889-904, 1989  
 A:Title: The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identificatio  
 A:Reference number: A33629; PMID:90072069; PMID:2574001  
 A:Accession: A33629  
 A:Molecule type: DNA  
 A:Residues: 1-373, 'V', 375-497 <KIM>  
 A:Cross-references: EMBL:M33388; NID:G181303; PIDN:AAA53500.1; PID:G181304  
 R:Manns, M.P.; Johnson, E.F.; Griffith, K.J.; Tan, E.M.; Sullivan, K.F.  
 J. Clin. Invest. 83, 1066-1072, 1989  
 A:Title: Major antigen of liver cytochrome P450 microsome autoantibodies in idiopathic autoimmu  
 A:Reference number: A30335; PMID:89155788; PMID:2466049  
 A:Accession: A30335  
 A:Molecule type: mRNA  
 A:Residues: 125-373, 'V', 375-485, 'T', 487-497 <MAN>  
 A:Cross-references: EMBL:M24499; NID:G522194; PIDN:AAA36403.1; PID:G522195  
 A:Genetics:  
 A:Gene: CYP2D6  
 A:Cross-references: GDB:132127; OMIM:124030  
 A:Map position: 22q13.1-22q13.1  
 A:Mutations: 60/3; 118/1; 169/1; 222/3; 281/3; 329/1; 391/3; 439/1  
 C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 monooxygenase  
 C:Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metal  
 F:302-465/Domain: cytochrome P450 homology <CYP>  
 F:443/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 3.0%; Score 117.5; DB 1; Length 497;  
 Best Local Similarity 21.7%; Pred. No. 1.2; Indels 153; Gaps 22;  
 Matches 97; Conservative 53; Mismatches 144

345 AVML-----QGLFHQHPQFQWPAQIRAVLAVGRGLVLDLFAESQFVYTRTAS 396  
 14 AIFLLVLDLHRRGRWAAAYP-----GLPLPGL-----GNLLHVD----- 50  
 397 FQGPFTWCLHNFGRN-----HGLFGLER-VNGGEAARL--FPNSTWV 439  
 51 FQNTPYCFDQLRRRFQGVFSLQIATPVTAVVNLGLAAVRELVHGEDTAPRPVPTIQL 110  
 440 GTGWAPPGI-----SQNEVYVSLMAELMKRDPVPLAAMTSPARRRYGVSHPD 489

Db 111 GFGPSQCVTLARGPANREORRPSVSTLRNLGLCK--KSLQWVTEBA----- 158  
 Qy 490 AGAAMRLILRSVYNGSGEACRGNHNSPLVRRPSLOMTSITWYNSDYFEAMRLILTSAPS 549  
 Db 159 -----CLCAAFANHSGRPF--RPNGLDKAV-----SNV-----AS 188  
 Qy 550 LATSAPRYD-----LIDLTRQAVQELVSLIYEASAYISKLASLIRAGVLAIE-- 601  
 Db 189 LTCGRFRFYDDPRFLRLIDLAQEBGLKESSGFLREVLNAVPLIHPAL--AGKYLROKA 246  
 Qy 602 LLPALDEVL-----ASDSRFLGSMLEQARAAVAEADPFYQNSRYQLTWGPE 652  
 Db 247 FLTQIDELLTEHRMTWDPAQPPRLTEAFLEMEKAK-GNPSSEFNENLR----- 296  
 Qy 653 GNILDYANKQLAGLVANYTTPRMRLFELALVDSVAQGIPOQHQPKXVPQLEQ----- 706  
 Db 297 ---IVADLFSAQWYTTSTLLAMGLIMILHPDQRRV---QGRIDVIGVRRPEMGDQ 350  
 Qy 707 -----AFVLSKORYPSPQPRGDTVDL 726  
 Db 351 AHMYTTAVIHEVQRF-----GDIVPL 372

## RESULT 7

D72593  
 hypothetical protein APEI213 - Aeropyrum pernix (strain KI)

C/Species: Aeropyrum pernix  
 C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C/Accession: D72593  
 R/Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hatake, Y.; Jin-no, K.; Takahawa, H.; Takamizawa, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999  
 A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix strain KI  
 A/Reference number: A12450; MUID:99310339; PMID:10382966  
 A/Accession: D72593  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1374 <KAM>  
 A/Cross-references: DDBJ:AP000061; NID:gs104821; PIDN:BA80202.1; PID:dl043988; PID:gs104821  
 A/Experimental source: strain KI  
 A/Genetics: APEI213

Query Match 3.0%; Score 117; DB 2; Length 1374;  
 Best Local Similarity 20.0%; Pred. No. 5.6;  
 Matches 163; Conservative 78; Mismatches 249; Indels 324; Gaps 37;

Qy 1 MEAVAVAAAVGVLLIAGG-----AAGDEAREAAVRA-----LVAR 38  
 Db 620 IDSTILETGVPHFIAAGNGGPGGLGTVPATARLAVAAATDMAVLSLIGPYLPLIAG 679  
 Qy 39 ILGPGPADPSV-SVERALAAKPGIDTYSIGGGGAARVRVGTGVAAGLHRYLDFC 97  
 Db 680 LGGYCDPRVFSARGSHGAKPKGI-----AATGFRYTTGRS 717  
 Qy 98 GCHVAMSGSQLRPLPLPAVPGELTEATPNRYRYQNVCTGYSFVMDMAREREIDMM 157  
 Db 718 LDH--YTGGRLD-PRAAALFGTSMATP-----M 744  
 Qy 158 ALINGINIAL-----AMSGGEAI-----MGRVYLAGLQAEINEFTGPAFLAMGMGLHT 209  
 Db 745 AAGAAALAIQALKESLVERLGLBEWMLVYALSTWTAQ-----WRG--LPMGMGN-- 793  
 Qy 210 WDGLPSPMHK-----QVLOHRYLDQMR-----SFGM-----TPV 241  
 Db 794 --GYDAGAIRLILGVQGLVYSATILEBAQAGVAPGYGIPALLVWAGSGVETPV 851  
 Qy 242 LPARAGHPAVATRPQVNTTKMSWGHNCSSCSFLAPEDPPIIIGSLFRLILIK 301  
 Db 852 DIVEGESPVALKVAEPRLLEVTVSS-----VEILSSPVG 887  
 Qy 302 EFGTDHIGADTFNEMQPSSEPSYLAATAVVAEAMTAVDTEAVWLLQGLFQHPQFV 361

Db 888 ESGYTMNNAASIDPSMLPQGP--VEVSLTLPYE-----VFDRCG--REERTSW 932  
 Qy 362 GPAQIRAVL-----GAVPRGRILVLDLFAESQPVYTRTASFQCPIMCMLHNF 410  
 Db 933 WEGYVYGVALLIYMDLDDGIDGDEFFILSIDKSSNVF-----RVF----- 976  
 Qy 411 GGNHGLFGLLEAVNGGPEARLFPNSTWYGTGMAPBGISQNEVYSLMAELGW--RKPD 467  
 Db 977 -----LSNQEALAGREA-----LGWS--EGVEEYVLRLLALAGLSMAGGTTA 1019  
 Qy 468 VEPDLA--WYTFEAPARVGVSHPDAGAAWRLILRSVYNGSGEAC----- 509  
 Db 1020 TVEIAAISVESA-----LPLPAIVEGSAVVAEAVNPPEPGYITGF 1063  
 Qy 510 -----RGHNRSPLV---RPSL-----QNTSTWYNSDYF 537  
 Db 1064 IVASTGVGEYRMPYTTLSPPRSHGVHVPQGLGDEYEEAMVGAFFYSWMYEDGD-- 1121  
 Qy 538 EAMRLILTSAPSLATSAPARYDLIDLTRQAVQELVSLIYEASAYISKLASLIRAGGV 597  
 Db 1122 --RIVPPELLPGGLAVARLHPVEDGREAVATNIDAY-----LYLNPISITSEAGEV 1173  
 Qy 598 L--AVELLPALDEVLASDSRF--LIGSWLE-----CARAAVSEADPFYQNSRYOL 646  
 Db 1174 VVGVEEMALASREASAPANRFDPITLGFWDQSGSGEYVLAAYNHA----- 1220  
 Qy 647 TLMGPEGNILDYANKQLAG-----LVANYTTP 673  
 Db 1221 ---PGRVLLYVRSVQYSGETARQPVVITLYHTP 1250

## RESULT 8

T35238  
 Probable secreted cellulase - Streptomyces coelicolor

C/Species: Streptomyces coelicolor  
 C/Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
 C/Accession: T35238  
 R/Seeger, K.D.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajadream, M.A. submitted to the EMBL Data Library, September 1998  
 A/Reference number: Z21572  
 A/Accession: T35238  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-973 <SEE>  
 A/Cross-references: EMBL:AL031515; PIDN:CAA20643.1; GSPDB:GN00070; SCOEDB:SC5C7.31C  
 A/Experimental source: strain A3(2)  
 A/Genetics: SCOEDB:SC5C7.31C

Query Match 2.9%; Score 115.5; DB 2; Length 973;  
 Best Local Similarity 18.4%; Pred. No. 4.4;  
 Matches 165; Conservative 93; Mismatches 293; Indels 347; Gaps 45;

Qy 1 MEAVAVAAAVGVLLIAG-----AGGAAGDEAREAAVRLVRLIGPGPADPSVSEVA 55  
 Db 225 MGSASDSTVGVVAAAGPTVASPQQLGVQGESGTYEV---KLSBQTAIVATVYTTSHA 280  
 Qy 56 LAAKPGIDTYSIGGG-----AARVAV-----RSGTGVAAGLHRYLDFCG 98  
 Db 281 -SGNTGL--TLTGASLFTPSNMDTAQRVVSADAGSGSAVFESTAPRGH----- 329  
 Qy 99 CHVAMSGSQLRPLPLPA-----VPGELTEA-----TPNRIRY-----QNV 136  
 Db 330 -KAAVTVTOLAARVDAFLELYGRTIDPANGYFSPGIPYHSVETLIVEAPDHGHTT 388  
 Qy 137 TQSYFVMDMAREREIDMVALNGINIALAMSGEAIWQRYVALGLQAEINEFTGP 196  
 Db 389 SEAYSTILM-----LQAMYGK-----TGD 408  
 Qy 197 AFLAMGRGNLHTWDGLPSPSHIKQVYQHRVLDQMSFGVTPVLPAPAGVPAVAT--R 255  
 Db 409 ---WSKFNN-----AMEIMETVMIPTHADQPTNSSYNASKP--ATYAPSLDTPN 452

QY 256 VEPQVNVTKMGSMGHFNCSYCSPLIAPEDPIFIIGSLFLRELKKEFTDHIYGA----- 311  
 DB 453 RFP-----APIDGTIVSGSDPIAGELMSAIGTIDVYGMHLQ 489  
 QY 312 ---DTFN-----EMOPSSSEPSYTL-----AAATTAVENTAVIDTEA-----VW 347  
 DB 490 DVNDITGYGNSPKCKAGPSDGTGSPSYINTFORAGQESVWEIVPOPTCDAFKYGKNGYLD 549  
 QY 348 LLOG-----WLFQHPQFQWPAQIRAVIAGVPRGRLLVLDLFAESQ-----PVYRTT 394  
 DB 550 LFTGDSYAKQKFTAPD-----ADARAYQAA-----YKADIMAGBOCKSDISATIDK 599  
 QY 395 ASFQGPFTWCMILHNGHGLFGALEAVNGEPAARLEPNSTWGTGMAREGISQNEVY 454  
 DB 600 AAKGQDYLRYAMEDKRYFKYKXG-----NCGVPSACFAGTKDSSH 638  
 QY 455 YSLMABLMGRKQDVPDLAAVMTSPARARYGVSHPDAGAWRLLLSYVNCSEACRGNR 514  
 DB 639 Y-----LISVYIANG-----GAVDTSHGMAWRI-----GSHHGGYQ 671  
 QY 515 SPLVR-----RPSLOQNTSIWYNRSD-VFEAWRLLLTS-----APSLAT 552  
 DB 672 NPLAAVALSTDDADLKPKSATGQSDMAKSLDRQVEFYRMLQSDGAIAGATNSMAGRYAT 731  
 QY 553 SPAPFRDLDTQAVQELVSLYEE-----ARSAVLSKELASLLRAGVILA 599  
 DB 732 PPA-----GTFPYGMVYDKPYHDPSPNQWFGFOAMSNERYAETYYQSG 777  
 QY 600 YEILPALDEVLASDSRFLGSMLEQARAAVSEAEADFEYQNSRYOLTV-----W 649  
 DB 778 -----DAGAAKAVLDKWDWMLSETTVNPDCTF-----RIPSTLQMSGPDTWNA 821  
 QY 650 ---GPBGN---LLDYANKQLAGIVA-----NYTPR-----KALFLEALVDV----- 686  
 DB 822 SSPGAGDLDHEVADYATND--VGVAAYAKXTLITYADRSGDTEAATAALLDGMENNQ 879  
 QY 687 -AAGIPFQOHQFQKVFQLEQAFVLQKQRYSPQPRADTVDLAK-----KIFLKYYPGW 738  
 DB 880 DALGIAVPEIRADYRNF--DDGIYVPSGMSGTMPNNDITVDASTFASISFYQDDPAM 935

RESULT 9  
 C87485  
 ComC/Rec2 family protein [imported] - Caulobacter crescentus  
 C/Species: Caulobacter crescentus  
 C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C/Accession: C87485  
 R/Nerman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
 B.; Taub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A/Title: Complete Genome Sequence of Caulobacter crescentus.  
 A/Reference number: A87249; MUID:21173698; PMID:11259647  
 A/Accession: C87485  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-725 <STO>  
 A/Cross-references: GB:AE005673; NID:g1342335; PIDN:AKX3879.1; CSPDB:GN00148  
 C/Genetics:  
 A/Genome: CCI904

Query Match 2.9%; Score 115; DB 2; Length 725;  
 Best Local Similarity 20.6%; Pred. No. 3.2;  
 Matches 134; Conservative 67; Mismatches 234; Indels 214; Gaps 29;

QY 1 MEAAVAAAVGVLILAGAGAGDEAR--EAAAVRALVAILGPAPADSVSEVSEALAA 58  
 DB 54 LEVAVPEALMGLTLLAGVITGLAAGARRMNAPELLIVLLGLIFAGAAGMAAATVRSERVA 113  
 QY 59 KPGL-----DTYSLGGG-----AARVRVSGTGVAAGAALHRYLRDFCGH 100  
 DB 114 APVMDGRIVRVGVVDVSPGAGGQRLILIAFVRVSG-----LSPEPTFRIRIRVITIGEN 169

QY 101 -VAMSGQLRL-----PRPLPAVGELETATENRYQNVCTQSYSE---VMMDMARWE 151  
 DB 170 PVKPKGQALGLRMLGPPPPAPG-----AYDFARDAMFD----- 205  
 QY 152 REIDMALKINLALAMSGEALWQVYVILGLTOAEINEFFGPPFLMGMKGNLHTWD 211  
 DB 206 -----SIGGVGPAIGDIGEVTLDQPP-LRLRLVNA-VNAFRMDLQRLIARAGPPISGGI 257  
 QY 212 GPLPSPWHIKQLYLQHRVLADQMSFGMTPLPAFAGHVEAVTVRVPQVNVTKMGSMGHF 271  
 DB 258 GAAMVICH--EAMISEQVETAMRASGLAHILISGLHM--AIVGCHVFV-VVRMGV----- 308  
 QY 272 NCYSYCSFLIAPEDPIFIIGSLFLRELKKEFTDHIYGDITNENOPSSSEPSYLAAT 331  
 DB 309 -AAMPMLALAPGKKAASAG-----LVSLGYLLVSGA-----PPAEPAALITASY 354  
 QY 332 TAV-----YEANTAVDEAVMLLOQM-----LFQ 355  
 DB 355 AFLAILEPDRRAITLHGALALALISILLKPEPAGEPQMSFAATVALALAESMPKPVRE 414  
 QY 356 HQPQW--GPAQIRAVIAGVPRGRLLVLDLFAESQPYTTRTASFOQPF-----IWCMLN 409  
 DB 415 LSTFWMIQGPQALATWL-AVSIASLVAGL-----ATAEPMAQHFRRVAVW----- 459  
 QY 410 FGNHGLFGALEAVNGEPAARLEPNSTWY-----GTGMADEGISQNEVYSLMAELG 462  
 DB 460 -----GLPAMLAIVPSLSFYIMFPAIGTLEEFGLSA-----PFLAVAG 499  
 QY 463 WRKQVVDLAAVMTSPARARYGVSHPDAG-----AMRLIASV----- 501  
 DB 500 WGIQVLMNVA-----GLPSEAHGQHIIVASAPQVLVAFLGLMVLCLMRRLMIGAPLA 555  
 QY 502 -----YNCSGEACRGHNSPVRPSLOMNTSIWYNR 533  
 DB 556 LAVALMREPTPPDAMIADGITAIVRSQNAVALIRTDKAFGEIWMARR 604

RESULT 10  
 T1293  
 aldehyde dehydrogenase homolog - Sphingomonas aromaticivorans plasmid pNL1  
 C/Species: Sphingomonas aromaticivorans  
 C/Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 04-Mar-2000  
 C/Accession: T1293  
 R/Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, C.W.;  
 submitted to the EMBL Data Library, July 1998  
 A/Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas aromac  
 A/Reference number: Z20992  
 A/Accession: T1293  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-479 <ROM>  
 A/Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378434; PIDN:AAD04017.1  
 C/Genetics:  
 A/Genome: plasmid pNL1  
 A/Note: nahr

Query Match 2.9%; Score 114; DB 2; Length 479;  
 Best Local Similarity 23.3%; Pred. No. 2.1;  
 Matches 112; Conservative 52; Mismatches 165; Indels 152; Gaps 24;

QY 222 QLYIQRVLDQMSFGMTPLVPAFAGHVEAVTVRVPQVNVTKMGSMGHF 278  
 DB 52 QAYIQHRDLPAVREGMT---AQAATIMEEDIKKADVLVDYDEIGSLPAKAGFETRAVS 107  
 QY 279 FLA-----PEDPIFIIGSLFLRELKKEFTDHIYGADTNNEMOPSSSEPS 325  
 DB 108 FLRAIAGVRRIRGETIPSDT--PGRFSMSLRGV-----GVVAGITPPFVFLIKIGIKIOS 160  
 QY 326 YLAANTTAVYEAAMAVDTEAVMLQGLFQHPQFQWPAQIRAVIAGVPRGRLLVLDLFA 385  
 DB 161 AMALATG---NAFVILPSEAPMTADLL---AKLMKEA-----GVPDGLFNV----- 201

```

QY 386 ESQPYTTRTASFGQGFPIWCMHNFGNGLFGALEAVNGGEAARL.FPNSTWVTGMA 444
DB 202 ---YV-----GNGAEIG--DVLTHGPKVASITFTGSSRGVKAIA 235
QY 445 PEGISONEVYVSLMELGWRKDPVPLDAMWTSFARARGVSHPDAGAMRLLSRYNC 504
DB 236 -EIAARNLKITYL-ELG-GKSPVYCA-----DADLDKAVNAALFSIMY 277
QY 505 GSEACGHNRSPLVRPRLQWNTSIWYNSDVFEAMRLLTSAPSLATSFAFRYDLDT 564
DB 278 QGVCMGASR-----IYERS-IFDQF---TXFAAATGRANSGDLRDP 318
QY 565 -----RQVQELVSLYEBARSATLSKELASLRAGVLAPELLALDEVLASDSRFL 617
DB 319 TMLGPIISERDRVRRIIDARS-----KGAAYLAG----- 350
QY 618 LGSWLEQARAAYSE---AEADFEQNSRYQTLWGPEGNI---LDYANKQLAGLVANY 671
DB 351 -GEMSGNSCAATILSGVTAMTVFEEETFGPTSLFPPTLEALELANNNEYGLSASIF 409
QY 672 T 672
DB 410 T 410

```

## RESULT 11

```

GB3405
Hypothetical protein PA1923 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83405
R:Stover, C.K.; Pham, X.O.; Ewitt, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Vann, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laddig, K.; Lim,
., Loty, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: G83405
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1281 <STO>
A:Cross-references: GB:AE004618; GB:AE004091; NID:g9947912; PIDN:AA05311.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1923

```

Query Match 2.9%; Score 113; DB 2; Length 1281;

Best Local Similarity 22.5%; Pred. No. 9.8; Mismatches 141; Indels 174; Gaps 26;

```

Matches 104; Conservative 44; Mismatches 141; Indels 174; Gaps 26;
QY 349 LOGWLFQHPQFQWGPQAIKAVLAGAVP-----RGRLLVLDLFAESQPYVTRTA---SFQ 399
DB 40 LAGMARAEAGVELRG---LRIGIGEARPGEMLDGNLLIDTPRPDRAQVEALGERIQ 96
QY 400 --QPTWCMHNFGNGLFGALEA-----VNGSP-----EAARLFENSTWVG 440
DB 97 GTQPMW-----RVGGPPEFGNLPALGRLVGYVANGEANLRRLFAVVRMAGLTVD 151
QY 441 TGAAPGISO-----NEVYSLMAE-LGMRKDPVPLDAAVNT-----SFAARYGYSH 487
DB 152 ALPAPQPLAQAGFYHPDAPPAVLAADYLAW-----GASRNASAPRAPIFLIPGAIAD 205
QY 488 PDGAAWRLLRSVYNGSGEACGHNRSPLVRPRLQWNTSIWYNSDVFEAMR----- 541
DB 206 AQTGAIDELRRS-----ERHGOAVL-----AVFPDSDP-EALRRKSFAGA 245
QY 542 -LLITSASLATSAPFRYDLDTLRQAVQELVSLYEE-----ARSAVLSKELASL 591
DB 246 DVQALVNLQNLQNGPARAEFLALD---VPVLTQTLGTRDGNADWLAAAGVAPRTAAAF 302
QY 592 -----LRAGVLAPELLALDEVL-----ASDSRELLGSW 621
DB 303 LGMPETWMSDPLVLSALENGEPKLMAG--QAEALIDKIDRLRLRRLPADKHLALMFW 360

```

```

QY 622 L-----EQARAAYSEADFEQNSRYQTLWGPEGNIIDY 658
DB 361 NHPEGKNVAASHLNPASLARLGEALRAAGYVAVNS-----ESALIDT 405
QY 659 ANKQAGLVANYTTPRRKRLLEALV-DVAQGIPEQO--HQFD 698
DB 406 AQRLIGQ-----YYRPQ---TLDALYRDGLAASLPLDAYLHWE 441

```

## RESULT 12

```

178557
N-methyl-D-aspartate receptor chain NMDAR2D-2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 21-Jan-2000
C:Accession: I78557; I58158; D45219
R:Monyer, H.; Burnashev, N.; Laurie, D.J.; Sakmann, B.; Seeburg, P.H.
Neuron 12, 529-540, 1994
A:Title: Developmental and regional expression in the rat brain and functional properti
A:Reference number: I58158; MUID:94206533; PMID:7512349
A:Accession: I78557
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1323 <RES>
A:Cross-references: GB:I31612; NID:g469068; PIDN:AAC37647.1; FID:g469069
A:Accession: I58158
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-66; 'V', '68-1323 <RE2>
A:Cross-references: GB:I31611; NID:g469066; PIDN:AAC37646.1; FID:g469067
R:Shi, T.; Moriyoshi, K.; Sugihara, H.; Sakurada, K.; Kadotani, H.; Yokoi, M.; Akazaw
J. Biol. Chem. 269, 2836-2843, 1993
A:Title: Molecular characterization of the family of the N-methyl-D-aspartate receptor
A:Reference number: A45219; MUID:93155102; PMID:8428958
A:Accession: D45219
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1265-1323 <ISH>
A:Experimental source: tRNA
A>Note: sequence extracted from NCBI backbone (NCBIP:124265)
C:Superfamily: N-methyl-D-aspartate receptor 2D; glutamate receptor homology
F,451-879/Domain: glutamate receptor homology <GRH>

```

Query Match 2.8%; Score 112; DB 2; Length 1323;

Best Local Similarity 22.9%; Pred. No. 12; Mismatches 157; Indels 152; Gaps 28;

```

Matches 106; Conservative 48; Mismatches 157; Indels 152; Gaps 28;
QY 162 INIALANGGEALWRYVYALGLTQALNEFFGPAFLANGKGNH-----TWGPP 215
DB 48 LNVAVFSSP-----AYAAEARL-----GPAVAAVRSPLDVRVALVINGSDP 93
QY 216 PSMHIKQLYQ-----HRVL--DQWSPGMPVLPAPAGHYBEAVTRVFPQVNV 263
DB 94 PS-----LVQLQDLDLGLRGVHGVFEDDSRAVAPILDFLSAQSLPIVAVHG----- 143
QY 264 KMGSGHFNCSYSCSPLADEDPITPGSLFIRELIKFGTDHYIYADTFNEMQ----- 318
DB 144 -----GAAVLTPKPK-----GSTFQ-----LGSSTEQQLQVIFEV 175
QY 319 -PSSPSRYAAATTA-----VYEAATVADTEAVTWLQGMFQHPQF-WGPQAIRVAV 372
DB 176 LEEYDWTSTFAVYTRAPGRAPFLSYIEVLTG--SLVGN--EHRGALLTDPGAEVLA 231
QY 373 VPRG-----RLVLDLFAESQPY--TRTASFGQPFIMWT--HNGGNHGLFGALEAVN 424
DB 232 QLRVSQAQRLIFC-AREEAPVFAAEKGLGPGVFMVFPQLAGGG-----S 282
QY 425 GSEBARLFPNSTWGTGMAPEGISONEVYVSLMELGWRKDPVPLDAMWTSFA----- 479
DB 283 GVPGEPLLEPGSSPLPAQ-----FAVASAGWMDLAKRVVAAVAAVAAQAQA 330
QY 480 -ARRYGSHPDGAAWRLLRSVYNGSGEACGHNRSPLVRPRLQ--MNTSIWYNRSD 535

```

Db 331 LIRDYGL-PEL-----GHDCRTQNRTH--RGSLSRYFNMTT-MQNRDY 371  
QY 536 VFEARMLLTSPAPSLATSPAFRYDLDTQAOVELSVLYEE 578  
Db 372 SFNEDGLVN-PSLV-----VISTRTDRTWVGSGWEOQ 404

## RESULT 13

H87316  
hypothetical protein CC0545 (imported) - Caulobacter crescentus  
C/Species: Caulobacter crescentus  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C/Accession: H87316  
R/Niemann, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolton  
N.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A/Title: Complete Genome Sequence of Caulobacter crescentus.  
A/Reference number: A87249; MUID:21173698; PMID:11259647  
A/Accession: H87316  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1027 <STO>  
A/Cross-references: GB:AE005673; NID:913421736; PIDN:AAK22532.1; GSPDB:GN00148  
C/Genetics:  
A/Genes: CC0545

Query Match 2.8%; Score 111; DB 2; Length 1027;  
Best Local Similarity 22.5%; Pred. No. 10;  
Matches 157; Conservative 47; Mismatches 217; Indels 278; Gaps 38;

QY 135 VCTOSYFVW-----WDMRNER-----IDMANGINLAL-----ANSQGAIV 175  
Db 68 VCDVSATGQWTLRASERAMDLKRLNDEGSLRKAIGMRDLDGADPATADLDLALGKAFS 127  
QY 176 QRVYVATGLQAEINERFTGPAPLAWGMGNLHTWDGFLPPSWH-1KQLYLQHRVLDQNR 234  
Db 128 PKA-MAMMERITEQEAAGBEL---RLATALEMSGALAPPAALLARLQAALNRDQAR 183  
QY 235 SFQMTPLVLPAGHVPNAVTRVPPQVNVTKGSGHFNCSYSCFLIAP-EDPI----- 287  
Db 184 R-GEILLADGFGQ-----EQAVETLTIDW-----INPDQDIERVAVLY 221  
QY 288 ---FPIIG-SLFLBELIKERCTDI-----YGADTNEM-----OPSSSE 324  
Db 222 VSGLPALGKSTLLHVAHQVAGQGVETVVRDLPRASLDVLQDLGISTEVAHQVASQLP 281  
QY 325 SYLAATTAAYEANT-----GAVPRGILL-VYDLPRAS----- 387  
Db 282 SAADLDQARLEATTITQETRLKGGSGIFPYALIGVQLQSGAGRLVWVLDLEVAR 341  
QY 349 LQGLVFOHQPOFQPPAQIRAVL-----GAVPRGILL-VYDLPRAS----- 387  
Db 342 GRGETHPRLQFLDMQDLAPAVLPLVVAAGRGSLPVDRAVETLNDLTDAAERL 401  
QY 388 -----QPVYTRT-ASFGQOPFIMCMLEHNGNHS-----LFGALEAVNG 426  
Db 402 LSADVPALMTQVYIALAEGDPLQRLATLAREGAGLTKGKGAITTYRSLERLGG 461  
QY 427 PEA-ARLP-----NSTMGVTGMAP-----EGISQNEVYISLMAEL-----GW 463  
Db 462 RHAKIARFIPLFRFNSSELLGAVAPVYLGRLSPGAAA-ALVAMELDQDTWLAGEBEG 520  
QY 464 RKDPVPD-----LAAMVTSFAARVGVSHPDAGAMRLLLSVYCSGEACRGHNSPLVR 519  
Db 521 -LTQOPDLRRLAGLTYDEHARRFGALHRR-AM-----FAE 556  
QY 520 RPSLOMTSIVNYSVFEARMLLTSA---PSLATSAPAR-VYLLDTQAOVELSVLY 575  
Db 557 RPELWAGAESLYHR---LQATRWAGPDALKGMNLAAGVAFOSADLEELP----- 602  
QY 576 YEARSATLSKELASLIRAGVLAEE---LLP-----ALDEV-----LASDR 615

Db 603 -EABADA-----LIRARGERSYEGRGDALTPSGVSPAADELMLISKGLAEASH 653  
QY 616 FL-----LGSMLEQAPAAA 629  
Db 654 FYGRVFPALDPSTSPADVALTLIMRTGRWTEARRLLA 692

## RESULT 14

F87285  
conserved hypothetical protein CC0295 (imported) - Caulobacter crescentus  
C/Species: Caulobacter crescentus  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C/Accession: F87285  
R/Niemann, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolton  
N.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A/Title: Complete Genome Sequence of Caulobacter crescentus.  
A/Reference number: A87249; MUID:21173698; PMID:11259647  
A/Accession: F87285  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-285 <STO>  
A/Cross-references: GB:AE005673; NID:913421436; PIDN:AAK22282.1; GSPDB:GN00148  
C/Genetics:  
A/Genes: CC0295

Query Match 2.8%; Score 109.5; DB 2; Length 285;  
Best Local Similarity 23.7%; Pred. No. 2.1;  
Matches 80; Conservative 36; Mismatches 90; Indels 131; Gaps 17;

QY 328 AATTAAYEANTADTEAVNLQGLVFOHQPOFQPPAQIRAVLAVPRGLVLDLFAES 387  
Db 3 AAPTAAVAGLTAASWLTATTCAMF-----ARRRLADGRVLSV----- 43  
QY 388 QPVYTRTASFOGQPF-----IMCMLEHNGNHLFGALAE----- 422  
Db 44 ---QSLAMGATYAAVLPIYIGVWLVLRFGA--GLRGVLAAPATGLVTPPEALVAS 96  
QY 422 ---VNGSPEAAR-----LFPNSTMGTCM-----AEGISQNEVYISLMAEL- 460  
Db 97 LIDQFAGCGDLAEALGRAPVCILHTAIVAGLAHHRRABASAKNTLLQALAE 156  
QY 461 -----LGRKDPVPDLAAMVTSFAARVGVSHPDAGAMRLLLSVYNS 505  
Db 157 RATSVDVAERLMWAGARRVFPV-DITAVEMFGAADVNVVHMGRGEG--LNRATLQ-S 211  
QY 506 GEA-----CRGNRSPLVTRPSLOMTSIVNYSVFEA-----WRLLTSPASLA 551  
Db 212 LEARLDPLRFARAH-RSALV-----NLSKVSQAQPLSDGSRLLTLASGAERV 257  
QY 552 TSPAFRYDLTD-LTRQAOVELSVLYFEARAGAVLSKE 587  
Db 258 TSRTYRDALIKRLGRQA-----SDSQAAHNSSE 285

## RESULT 15

T47641  
hypothetical protein T15C9.20 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C/Accession: T47641  
R/Mewes, H.W.; Rudd, S.; Lemcke, K.; Meyer, K.F.X.  
submitted to the Protein Sequence Database, April 2000  
A/Reference number: Z24470  
A/Accession: T47641  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-814 <NEW>  
A/Cross-references: EMBL:AL122970  
A/Experimental source: cultivar Columbia; BAC clone T15C9  
C/Genetics:  
A/Map position: 3

A:introns: 28/2; 51/3; 252/3; 312/3; 350/3; 370/2; 392/3; 410/1; 441/1; 469/1; 604/3; 62  
A>Note: T1509.20

Query Match 2.8%; Score 109.5; DB 2; Length 814;

Best Local Similarity 20.6%; Pred. No. 9.3;  
Matches 113; Conservative 69; Mismatches 192; Indels 175; Gaps 26;

```
QY 150 WEREIDWALNGINIALAMSGOEAIWQRYVALGLTQAEINEFTGPALWGRMGVLT 209
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 233 WKEBLEVILIRGVPMAL-----RGELMQAF--VGVRKRCKDYQ-----NLTA 274

QY 210 WDGFLPESMHIKQYLQHRVLDDKRSFCMTFVLDPAFAGHVEATRYVEPQVNVTKGSMG 269
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 ADGSVNT---TEQEMQH-VDDKSGSTISIAVEKMKQIKNDPRTFP-----G 320

QY 270 HFNCSYSGSFLLAPDPIFFPIGSLFLRELKKEFGTDHIYGADTFNEMQPSSEPSYLA 329
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 H-----PALDDGRVALRRLTLAYARHN-----PSVGYC 349

QY 330 ATTAYEAMTAV---DTEAVVLLQGMLEQHQPQFWGPAQIRAVLGAVRGRLLVLDLFAE 366
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 QAMNFFALLLIMPEENAFMALIGLIDYFNNGYSEMTESQV---DQLVLEELVRE 404

QY 387 SQPVYTRTASFGQGFPIWCMLEHFGNHLFGALEAVNGSPEAARLFPNSTWGTGMAP- 445
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 405 RFPKLVIHLDYLGQVAVMT-----GPMFLIFMN-----MLPW 438

QY 446 EGISONEVYVYSLMELGWRKDPVPDLAAMVTSFAARR-YG--VSHPDAGAMRL---LL 498
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 439 ESVLR---VMDVLLFEGTR-----VMLFTALIMELYGPALVTTKDGADAVTLQSLT 489

QY 499 RSVNCSG--EACGHRSPVLRPSLQWN-----TSIWTNRSDVEAMR----- 541
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 480 GSTPSSQVLTLACMGYQNVHEIRLQELRSKRPAVIAALEERSKGLQAMRDSKGLASKL 549

QY 542 -----LLTTSAPSLATSPAPRYD-----LLDLT-----RQAVQELVS 573
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 550 YNFKODPKSVLVDSKASLSNGSLSRSESGSSNADEVLSLTGDEVDVSVQDLQAQAEICK 609

QY 574 LYFEARBARVLSKELASLLRAGVLAAYELLPALDEVLASDSRFLGSMLEQ--ARAAVS 631
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 610 L-LBEKRSALLRAE-----ELEIALMEIYKEDNRROLSAKVEQLBOEMAEVQ 655

QY 632 EAADFYEQ 640
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 656 RLSDKQEQ 664
```

Search completed: February 13, 2004, 16:16:27  
Job time : 37 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2004, 13:34:59 ; Search time 19 Seconds

(without alignments)  
1838.992 Million cell updates/sec

Title: US-09-836-613-2

Perfect score: 3939  
Sequence: 1 MEAVAVAAVGVLLIAGAG.....VDLAKKIFLKYPGWAGSW 743

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3939	100.0	743	1 ANAG_HUMAN	P54802 homo sapien
2	123.5	3.1	497	1 CPDH_MACFA	Q29488 macaca fasc
3	119	3.0	3649	1 ACVS_NOCILA	P27743 nocardia la
4	117.5	3.0	497	1 CPD6_HUMAN	P10635 homo sapien
5	110.5	2.8	1029	1 YK95_RHIME	Q52999 rhizobium m
6	109.5	2.8	1377	1 RHSA_ECOLI	P16916 escherichia
7	109.5	2.8	1411	1 RH8B_ECOLI	P16917 escherichia
8	108	2.7	726	1 CATA_SALTY	P17750 salmoneilla
9	104.5	2.7	726	1 CATA_ECOLI	P13029 escherichia
10	103	2.6	722	1 P85B_RAT	Q63788 rattus norv
11	101	2.6	698	1 YMCA_ECOLI	P75882 escherichia
12	101	2.6	3519	1 OL56_STRAT	Q07017 streptomyce
13	100	2.5	726	1 CATA_SALTY	Q82303 salmoneilla
14	100	2.5	851	1 ENV_HVZDI	P17755 human immun
15	100	2.5	1323	1 NME4_MOUSE	Q03391 mus musculu
16	100	2.5	1336	1 NME4_HUMAN	Q15399 homo sapien
17	99.5	2.5	1092	1 LIFR_MOUSE	P42703 mus musculu
18	99	2.5	624	1 SIR_SYNP7	P30008 synecococc
19	99	2.5	1330	1 UGSA_SOLITU	Q43846 solanum tub
20	99	2.5	1587	1 LMG3_HUMAN	Q9466 homo sapien
21	98.5	2.5	512	1 NUON_RHOCA	P50974 rhodospacer
22	98.5	2.5	1774	1 MSAS_PENPA	P23267 penicillium
23	98	2.5	860	1 ENV_HVZBE	P18094 human immun
24	98	2.5	973	1 VP18_HUMAN	Q92253 homo sapien
25	98	2.5	1323	1 NME4_MOUSE	Q62845 rattus norv
26	98	2.5	2067	1 BIME_EMENT	P33144 bacillus ha
27	97.5	2.5	497	1 GLPR_BACND	Q9466 bacillus ha
28	97.5	2.5	693	1 UL47_HSV11	P10231 herpes simp
29	97.5	2.5	786	1 AAC_ACTUT	P29958 actinoplan
30	97.5	2.5	842	1 VGLH_HSVBC	P27599 bovine hept
31	97.5	2.5	1155	1 C2TA_MOUSE	P79621 mus musculu
32	97	2.5	1094	1 EMBC_MYCTU	P72029 mycobacteri
33	97	2.5	2594	1 TLBS_DROVI	P20806 drosophila

34	96.5	2.4	856	1 ENV_HVIMN	P05877 human immun
35	96	2.4	708	1 NICA_MOUSE	P57716 mus musculu
36	96	2.4	2314	1 PTP2_HUMAN	P23471 homo sapien
37	95.5	2.4	330	1 RCEN_ROSDE	P26279 roseobacter
38	95.5	2.4	914	1 PERT_RAT	P14650 rattus norv
39	95.5	2.4	970	1 ANDI_MOUSE	P59328 mus musculu
40	95	2.4	344	1 LICI_BURCE	P22089 burholderi
41	95	2.4	636	1 GIDA_BRUME	Q8Y35 bruceella me
42	94.5	2.4	664	1 UL47_HSV1F	P08313 herpes simp
43	94.5	2.4	761	1 CTPA_MYCTU	Q10876 mycobacteri
44	94.5	2.4	866	1 NASA_KLEPN	Q06457 klebsiella
45	94	2.4	500	1 CPD6_CAVPO	Q64403 cavia porce

## ALIGNMENTS

### RESULT 1

ANAG\_HUMAN STANDARD; PRT; 743 AA.

AC P54802; 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Alpha-N-acetylglucosaminidase precursor (EC 3.2.1.50) (N-acetyl)-alpha-glucosaminidase (NAG).  
CN NAGLU OR UHSDI.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX MEDLINE=96324097; PubMed=8650226;  
RA Zhao H.G., Li H.H., Bach G., Schmidtchen A., Neufeld E.F.;  
RT "The molecular basis of Sanfilippo syndrome type B.";  
RL Proc. Natl. Acad. Sci. U.S.A. 93:6101-6105(1996).  
[2]  
RN SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.  
RX MEDLINE=96372812; PubMed=8776591;  
RA Weber B., Blanch L., Clements P.R., Scott H.S., Hopwood J.J.;  
RT "Cloning and expression of the gene involved in Sanfilippo B syndrome (mucopolysaccharidosis II B).";  
RL Hum. Mol. Genet. 5:771-777(1996).  
[3]  
RN SEQUENCE FROM N.A.  
RA Zhao Z., Yazdani A., Shen Y., Sun Z.S., Bailey J., Caskey C.T.,  
RX Lee C.C.;  
RT Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.  
[4]  
RN VARIANTS MPS-IIIB HIS-92; SER-115; CYS-140; LYS-153; LEU-358; VAL-664 AND ARG-682.  
RX MEDLINE=98107938; PubMed=9443678;  
RA Schmidtchen A., Greenberg D., Zhao H.G., Li H.H., Huang Y., Tieu P.,  
RA Zhao H.-Z., Cheng S., Zhao Z., Whitely C.B., di Natale P.,  
RX Neufeld E.F.;  
RT "NAGLU mutations underlying Sanfilippo syndrome type B.";  
RL Am. J. Hum. Genet. 62:64-69(1998).  
[5]  
RN VARIANTS MPS-IIIB.  
RX MEDLINE=99133861; PubMed=9950362;  
RA Bunge S., Knigge A., Steglich C., Kleijer W.J., van Diggelen O.P.,  
RA Beck M., Gal A.;  
RT "Mucopolysaccharidosis type IIIB (Sanfilippo B): identification of 18 novel alpha-N-acetylglucosaminidase gene mutations.";  
RL J. Med. Genet. 36:28-31(1999).  
[6]  
RN FUNCTION INVOLVED IN THE DEGRADATION OF HEPARAN SULFATE.  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing N-acetyl-D-glucosamine residues in N-acetyl-alpha-D-glucosaminides.  
CC -1- SUBUNIT: Monomer and homodimer.  
CC -1- TISSUE SPECIFICITY: LIVER, OVARY, PERIPHERAL BLOOD LEUCOCYTES, TESTIS, PROSTATE, SPLEEN, COLON, LUNG, PLACENTA AND KIDNEY.

CC -1- DISEASE: DEFECTS IN NAGLU ARE THE CAUSE OF MUCOPOLYSACCHARIDOSIS  
 CC TYPE IIIB (MPS-IIIB) (ALSO KNOWN AS SANFILIPPO B SYNDROME). AN  
 CC AUTOSOMAL RECESSIVE DISORDER WHOSE CLINICAL FEATURES ARE SEVERE  
 CC MENTAL RETARDATION BUT MILD SOMATIC MANIFESTATIONS IN CHILDHOOD,  
 CC AND DEATH IN THE SECOND DECADE. BIOCHEMICALLY, THIS DISEASE IS  
 CC CHARACTERIZED BY UNDEGRADED OR PARTIALLY DEGRADED HEPARAN SULFATE  
 CC WHICH ACCUMULATES IN LYSOSOMES AND IS EXCRETED IN URINE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; U43573; AAC05012.1; -;  
 DR EMBL; U43573; AAC05013.1; -;  
 DR EMBL; U40846; AAB06188.1; -;  
 DR EMBL; U78464; AAB36604.1; -;  
 DR PIR; G02270; G02270.  
 DR Genew; HGNC; 7632; NAGLU.  
 DR MIM; 252320; -;  
 DR GO; GO:0005764; C:lysosome; TAS.  
 DR GO; GO:0004561; F:alpha-N-acetylglucosaminidase activity; TAS.  
 DR GO; GO:0007399; P:neurogenesis; TAS.  
 DR Pfam; PF05089; NAGLU; 1.  
 DR Hydrolase; Glycosidase; Glycoprotein; Mucopolysaccharidosis;  
 KM Signal; Disease mutation; Polymorphism.  
 FT CHAIN 1 23  
 FT 24 743  
 FT 59 743  
 FT 68 71  
 FT 84 87  
 FT 134 134  
 FT 261 261  
 FT 272 272  
 FT 435 435  
 FT 503 503  
 FT 526 526  
 FT 532 532  
 FT 79 79  
 FT 92 92  
 FT 100 100  
 FT 115 115  
 FT 140 140  
 FT 142 142  
 FT 153 153  
 FT 243 243  
 FT 277 277  
 FT 280 280  
 FT 292 292  
 FT 358 358  
 FT 452 452  
 FT 482 482  
 FT 561 561

FT VARIANT 565 565 R -> Q (IN MPS-IIIB).  
 FT /FTid=VAR\_008989.  
 FT R -> H (IN MPS-IIIB).  
 FT /FTid=VAR\_005012.  
 FT A -> V (IN MPS-IIIB).  
 FT /FTid=VAR\_005013.  
 FT R -> H (IN MPS-IIIB).  
 FT /FTid=VAR\_005014.  
 FT L -> R (IN MPS-IIIB).  
 FT /FTid=VAR\_005015.  
 FT E -> K (IN MPS-IIIB).  
 FT /FTid=VAR\_008990.  
 FT G -> R.  
 FT /FTid=VAR\_008991.  
 FT A -> L (IN REF. 2; AA SEQUENCE).  
 FT S -> L (IN REF. 2; AA SEQUENCE).  
 FT S -> L (IN REF. 2; AA SEQUENCE).  
 SQ SEQUENCE 743 AA; 82166 MW; 6D8D6A42C7BA6083 CRC64;  
 Query Match 100.0%; Score 3939; DB 1; Length 743;  
 Best Local Similarity 100.0%; Pred. No. 1,7e-277;  
 Matches 743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MEAAVAAAVGVTLLAGAGAGDEAREAAVRLVARTLGGPADPVSVERALAKP 60  
 DB 1 MEAAVAAAVGVTLLAGAGAGDEAREAAVRLVARTLGGPADPVSVERALAKP 60  
 QY 61 GLDTYSLGGGGAARVRVRSVTGVAAGLHRYLDFGCGHVAWSGQLRPLRPVAVGE 120  
 DB 61 GLDTYSLGGGGAARVRVRSVTGVAAGLHRYLDFGCGHVAWSGQLRPLRPVAVGE 120  
 QY 121 LTEATPNRYRYQVNCVCSYFVWMDVAREERLDMALNGINLALMSGEALWQRYL 180  
 DB 121 LTEATPNRYRYQVNCVCSYFVWMDVAREERLDMALNGINLALMSGEALWQRYL 180  
 QY 181 ALGLTQAEINEFFGPAPFLAMGRMNLHTWGPPLPSPMHIKQYLQHRVLDQMSFCWTP 240  
 DB 181 ALGLTQAEINEFFGPAPFLAMGRMNLHTWGPPLPSPMHIKQYLQHRVLDQMSFCWTP 240  
 QY 241 VLPAPAGVPAVTRVPQVNTVMGSGHFNCSYSCSFLIAPDPPIPIIGSLFLELI 300  
 DB 241 VLPAPAGVPAVTRVPQVNTVMGSGHFNCSYSCSFLIAPDPPIPIIGSLFLELI 300  
 QY 301 KEFGDHIYAGDTENQMPSESPSYLAATYAYEAMTAVDTAAWLLQWLFQHPQF 360  
 DB 301 KEFGDHIYAGDTENQMPSESPSYLAATYAYEAMTAVDTAAWLLQWLFQHPQF 360  
 QY 361 WGPQIRAVLGAVERGRLLVLDLFAESQPYTRTASFQGPFIWCMHNFGNHLFGAL 420  
 DB 361 WGPQIRAVLGAVERGRLLVLDLFAESQPYTRTASFQGPFIWCMHNFGNHLFGAL 420  
 QY 421 EAVNGPBAALFENSTWVGMAPEGISONEVYSIMAEIGMRKDPVPLLAAVTSFPA 480  
 DB 421 EAVNGPBAALFENSTWVGMAPEGISONEVYSIMAEIGMRKDPVPLLAAVTSFPA 480  
 QY 481 RRYGSHPDAGAAWRLILRSYVNCGEACRGNHNSPLVRBPSLOMNTSIWNRSDVEAW 540  
 DB 481 RRYGSHPDAGAAWRLILRSYVNCGEACRGNHNSPLVRBPSLOMNTSIWNRSDVEAW 540  
 QY 541 RLITTSAPSLATSAFRYDLDLTLROAVOEIVSYIYEASAYISKELASILRAGGLAY 600  
 DB 541 RLITTSAPSLATSAFRYDLDLTLROAVOEIVSYIYEASAYISKELASILRAGGLAY 600  
 QY 601 ELPLALDEVILASDSRFLIGSWLEBARAAVSEAEADFEQNSRYQLTLWSPGEGNILDYAN 660  
 DB 601 ELPLALDEVILASDSRFLIGSWLEBARAAVSEAEADFEQNSRYQLTLWSPGEGNILDYAN 660  
 QY 661 KQLAGLVANVTTPKWRFLFLEALVDSVAQGIPOQHOPDKVVFQLEQAFVLSKQYPSQPR 720  
 DB 661 KQLAGLVANVTTPKWRFLFLEALVDSVAQGIPOQHOPDKVVFQLEQAFVLSKQYPSQPR 720  
 QY 721 GDTYDLAKKIFLKTYPGWVAGSW 743  
 DB 721 GDTYDLAKKIFLKTYPGWVAGSW 743

```

RESULT 2
CPDH_MACFA
ID_CPDH_MACFA STANDARD; PRT; 497 AA.
AC 029488;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 2D17 (EC 1.14.14.1) (CYP2D17).
GN CYP2D17
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OC NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Lawton M.P., Laddison K.J., Speirs A.A., Mankowski D.C., Tweedie D.J.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -1 CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -1 SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC similarity).
CC -1 SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U38218; AAA79722.1; -
DR PIR; G02938; G02938.
DR HSSP; P00179; 1DT6.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450_1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450_1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum.
FT METAL 443 443 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SO SEQUENCE 497 AA; 56010 MW; 3594AA88F04E58B1 CRC64;

Query Match 3.1%; Score 123.5; DB 1; Length 497;
Best Local Similarity 22.9%; Pred. No. 0.21;
Matches 104; Conservative 47; Mismatches 140; Indels 163; Gaps 22;

CY 343 TEAVWL-----OGWLFQHOPOFMGPAQIRAVLGVPRGRLLVLDLFABSQPVYTRT 394
DB 12 TVALFLVLVDLMHRRQRWAAARYPP--GDLPLPGL-----GILLHYD----- 50
CY 395 ASFOGQPIWCMILNFGGN-----HGLFGALIA-VNGSPEAART--FPNST 437
DB 51 --FKMTPYCFDPLRRRFGNVFSLQAMTPVYVNLGLAAVEALVTCEDPTADRPVYINQ 108
CY 438 MVGTGMAGEPT-----SQHEVYYSIMAEIGMKRDPVLDLAAVTSPFAARYGVSH 487
DB 109 VLGGRPSQGVFLARYGPAMREQRFRSVSTIRNLGLGK--KSLEQWTEBAACLAFT 165
CY 488 PDAGAAAR--LLRSVYNSGEGACRGHNSPVLVRBSLQWNTSIWYNSDPVEAMWLL 544
DB 166 DQAGPFRFNSLLDKAVSN-----VIALTYGR----- 153
CY 545 TSAASLATSAPFRD-----LLDITROAVQELVSLYEEKRSAYLSKELAS---LIR-- 593
DB 194 -----RFEYDDPRFLRLFDLTLEALKE-----ESGFLRELVNALPILLRIP 234
CY 594 --AGVLAVE--LIPALDEVL-----ASDRFLLGSWLEPARRAAVSEADPEYEQ 640

```

```

DB 235 GLAGKVLRSQAKFLTQDELLETHRMTPDPAQPRDLTEAFLAEMEKAK-GNPSSFNEE 293
CY 641 NSRYQLTLMGPEGNLTIDVANKQLAGVANYYPERRRLFLFEALVDVSAQIPQCHQFDKN 700
DB 294 NLR-----MVVADLFSGAVTTSTTLANGLLMLHPVQRRV---QQEIDV 338
CY 701 VFQLSQAFVLSKQRYV-----SQPRGTVLDL 726
DB 339 IGVRRPEMGDQARMFYTTAVTIVHEVQRFQDVIPL 372

RESULT 3
ACVS_NOCCLA
ID ACVS_NOCCLA STANDARD; PRT; 3649 AA.
AC P27743;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE N-(5-amino-5-carboxypentanoyl)-L-cysteine]-D-valine synthase
DE (EC 6.3.2.26) (Delta-(L-alpha-aminoadipyl)-L-cysteine]-D-valine
DE synthetase) (ACVS).
GN PCBAB.
OS Noccardia lactamdurans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardiaceae; Pseudonocardiaceae; Amycolatops.
OC NCBI_TaxID=1913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VAR LC 411;
RX MEDLINE=92065808; PubMed=1956290;
RA Coque J.J.R., Martin J.F., Calzada J.G., Liras P.;
RT "The cephamycin biosynthetic genes pcbAB, encoding a large
RT multidomain peptide synthetase, and pcbC of Noccardia lactamdurans are
RT clustered together in an organization different from the same genes
RT in Acremonium chrysogenum and Penicillium chrysogenum."
RT Mol. Microbiol. 5:1125-1133(1991).
RL -1 FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE
RL ACV ARE ACTIVATED AS AMINOACYL-ADENYLYATES WITH PEPTIDE BONDS
RL FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER
RL INTERMEDIATES.
CC -1 CATALYTIC ACTIVITY: L-2-aminohexanoate + L-cysteine + L-valine
CC + 3 ATP = N-[L-5-amino-5-carboxypentanoyl]-L-cysteine]-D-valine +
CC 3 AMP + 3 diphosphate.
CC -1 COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETINES.
CC -1 PATHWAY: Biosynthesis of penicillin and cephalosporin, first step.
CC -1 SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -1 SIMILARITY: Contains 3 acyl carrier domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57310; CAA40561.1; -
DR PIR; S18268; S18268.
DR HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR006163; PP_bind.
DR InterPro; IPR006162; Pantate_attach.
DR InterPro; IPR000379; Ser_estr_site.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00501; AMP-binding; 3.
DR Pfam; PF00568; Condensation; 3.
DR Pfam; PF00550; PP-binding; 3.
DR Pfam; PF00575; Thioesterase; 1.
DR PROSITE; PS00455; AMP BINDING; 1.
DR PROSITE; PS00455; AMP BINDING; 1.
DR PROSITE; PS00075; ACP_DOMAIN; 3.

```

KM Ligase; Antibiotic biosynthesis; Multifunctional enzyme;  
 KM Repeat; Phosphopantetheine.  
 FT REPEAT 401 861 DOMAIN 1 (ADDPATE-ACTIVATING).  
 FT REPEAT 1014 1937 DOMAIN 2 (CYSTEINE-ACTIVATING).  
 FT REPEAT 2079 2985 DOMAIN 3 (VALINE-ACTIVATING).  
 FT DOMAIN 788 857 ACYL CARRIER (ACP) 1.  
 FT DOMAIN 1864 1933 ACYL CARRIER (ACP) 2.  
 FT DOMAIN 2910 2981 ACYL CARRIER (ACP) 3.  
 FT BINDING 820 820 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 FT BINDING 1896 1896 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 FT BINDING 2944 2944 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 FT ACT SITE 3502 3502 THIOESTERASE (BY SIMILARITY).  
 SQ SEQUENCE 3649 AA; 404079 MW; 6FD095704F858EB8 CRC64;  
 Query Match 3.0%; Score 119; DB 1; Length 3649;  
 Best Local Similarity 21.1%; Pred. No. 6.3; Indels 306; Gaps 37;  
 Matches 174; Conservative 63; Mismatches 283;  
 16 AGAGGAAGDEAREMAAVALV-----ARLLGFPADPSVSVERALAAKPGLDYSLGG 69  
 1988 AALPGALGLVRRHPALRTLLKTDQGVRRQYIPADVVLEVP----- 2031  
 70 GGAARVAVRSTGVAAGAHLRYLRDPCGCHVAMSGQALRPPLPAVPGELTEATPNRY 129  
 2032 -----STVYDSFAELDEVLT-----RAGVYRLHEELPI-----RAEAFDGD 2070  
 130 RYQYQVCTQSYSFVWMDWAREREI-----DMALNGINIALAMSGGEATVQRY 179  
 2071 EYLSVAVHSCFGWMDIFRELAALLDGVEADIGALRG-----YGEFAVWQRY 2124  
 180 LALGLTQAEINERPTGTAFLAMGKMLNT-----WGQPLPSNHR-----QYLCORVID 231  
 2125 LT-GKRLAALTEWT-----GALGFEETIALPLDHPRPDYGRLEHEFELDERTE 2176  
 232 QMSFGMT----- 239  
 2177 ALRLATATAVSLVSLGAMCLMANTYQGHDLVVTGPRANKGREPRAVGFANLLA 2236  
 240 -----PVLPAFAGHPAATRVPPQVNTKMGKMGHNCYSQSFLLAPDPI-FPI 290  
 2237 LRYAVDPAATLPAYRVSVEAV-VAAQVH-----GELPFOVYKELVEEDPSRHP 2287  
 291 IGSFLRELLKEFTDIYGAFTENEMQPSSEPSYLAATVAEAMTADTEAWMLQ 350  
 2288 LQNLFTLQNV-----SDH-----TSALTGYQPS-----GQTTTKFD-LSMTMETATGLA 2333  
 351 GWT-----FOHQPFQWGPQAQIRAVLGAVPGRLLVIDLFAESQ-PY 391  
 2334 GNLTYAASLPDDTSAGFIATFKV-----LAEFASAACTPIAQLTALDEPQALPDA 2388  
 392 TRTASFOGQPIWCMANFGNHLFGALEAVNGSPRAALFNSWTGVS---MAPEG 448  
 2389 TRRARPPGCP-----GRCTRLP-----BEVAATWPRVAVVGDVRLTYREL 2430  
 449 SQ-----NEVVYSIMAEIGMKDP-----VPDLAAVMTSFAARVGVSHPDAGAA 493  
 2431 NERANRLAHLHRSVAEPRADELIALVLDKSELTVAILAVKGAAYMPIDPSYDRIA 2490  
 494 W-----RLLRSVNCSGEGACGHNRSPLVRPSLQMTSTWNRSDVFEAMRLITS 546  
 2491 FMSDGTAKVL-----AGEAGHSRYRG-----LTSGDVLDLBDQLDITG 2529  
 547 APSLATSPAFRYDLDLTRCAVQELVSLYEEARSAYLSKELASLAFAGVLAVELPAL 606  
 2530 EP-AENP-----VTEETSTELAYAIY---TSGTGPKAVLVHSGV--- 2567  
 607 DEVLASDSRFLGSLWLEQAPAAVSEADPYEQNSRYQLTMPENIL----- 656  
 2568 DSFPAQUSGRYFGSPDSAEAVLFL-----ANYVDFSVQALSYLGSHKLLVPPPSAAD 2624  
 657 -----DYANKOLAGLVANYTTRMFLFLALVDVSAQGIFFQOHFD 698  
 2625 PAFELANRE--GL-----SYLSTGTPQVVERFD 2650

RESULT 4  
 CPD6\_HUMAN STANDARD; PRT; 497 AA.  
 ID CPD6\_HUMAN  
 AC P10635; Q16752;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Cytochrome P450 2D6 (EC 1.14.14.1) (CYP1D6) (P450-DB1) (Debrisoquine 4-hydroxylase).  
 GN CYP2D6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88314109; PubMed=3410476;  
 RA Gonzalez F.J., Vilbois F., Hardwick J.P., McBride O.W., Nebert D.W., Gelboin H.V., Meyer U.A.;  
 RT "Human debrisoquine 4-hydroxylase (P4501D1): cDNA and deduced amino acid sequence and assignment of the CYP2D locus to chromosome 22.";  
 RL Genomics 2:174-179(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=8812614; PubMed=3123997;  
 RA Gonzalez F.J., Skoda R.C., Kimura S., Umeno M., Zanger U.M., Nebert D.W., Gelboin H.V., Hardwick J.P., Meyer U.A.;  
 RT "Characterization of the common genetic defect in humans deficient in debrisoquine metabolism.";  
 RL Nature 331:442-446(1988).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90072069; PubMed=2574001;  
 RA Kimura S., Umeno M., Skoda R.C., Meyer U.A., Gonzalez F.J.;  
 RT "The human debrisoquine 4-hydroxylase (CYP2D) locus: sequence and identification of the polymorphic CYP2D6 gene, a related gene, and a pseudogene.";  
 RL Am. J. Hum. Genet. 45:889-904(1989).  
 RN [4]  
 RP VARIANT LVS-281 DEL (CYP2D6\*9).  
 RX MEDLINE=93244880; PubMed=1844820;  
 RA Tyndale R., Aoyama T., Broly F., Matsumaga T., Inaba T., Kalow W., Gelboin H.V., Meyer U.A., Gonzalez F.J.;  
 RT "Identification of a new variant CYP2D6 allele lacking the codon encoding Lys-281: possible association with the poor metabolizer phenotype.";  
 RL Pharmacogenetics 1:26-32(1991).  
 RN [5]  
 RP VARIANTS SER-34 AND THR-486 (CYP2D6\*10).  
 RX MEDLINE=94115362; PubMed=8287064;  
 RA Yokota H., Tamura S., Puruya H., Kimura S., Watanabe M., Kanazawa I., Kondo I., Gonzalez F.J.;  
 RT "Evidence for a new variant CYP2D6 allele CYP2D6U in a Japanese population associated with lower in vivo rates of sparteine metabolism.";  
 RL Pharmacogenetics 3:256-263(1993).  
 RN [6]  
 RP VARIANT PRO-324 (CYP2D6\*7).  
 RX MEDLINE=95147995; PubMed=7845481;  
 RA Evert B., Griese B.U., Bicheldaum M.;  
 RT "A missense mutation in exon 6 of the CYP2D6 gene leading to a histidine 324 to proline exchange is associated with the poor metabolizer phenotype of sparteine.";  
 RL Naunyn-Schmiedeberg Arch. Pharmacol. 350:434-439(1994).  
 RN [7]  
 RP VARIANT GLU-212 (CYP2D6\*6B/6C).  
 RX MEDLINE=95172594; PubMed=7868129;  
 RA Daly A.K., Leathart J.B., London S.J., Idle J.R.;  
 RT "An inactive cytochrome P450 CYP2D6 allele containing a deletion and a base substitution.";



```

Db 159 -----CLCAFANSGRPF--RPNGLDKAV---SNVI-----AS 188
Cy 550 LATSPARVD-----LIDLTRCAVQELVSLYEARSAYLSKELASLIPAGVLA-- 601
Db 189 LTCRRREYDDPRFLRLILDLAQEGAKSSGFLREVLNAVPLILHPL--AGKTLRQKA 246
Cy 602 LLPALDEVL-----ASDPSFLGSMLEQAPAAAVSEADPFYQNSRYQLTLWGPE 652
Db 247 FLTQIDELITLHRTWDPACPPRDLTEAFLEMEKAK-GNPESEFNDENLR----- 296
Cy 653 GNILDYANKOLAGVANYVTYPRMFLFEALVDVAQGIPIFOQDFDKVFLQLEO----- 706
Db 297 ---IVADLFSAQWTTSTTLAKGLMLTHPDVQKRV---QGEIDVITQVRRPMGQ 350
Cy 707 -----AFVLSKORYPSQPRGDTVDL 726
Db 351 AHMPYTAIVHEVQRF-----GDIYPL 372

RESULT 5
ID YK95_RHIME STANDARD; PRT; 1029 AA.
AC Q52999;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein R02095.
GN R02095 OR SKC01491.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_Taxid=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Boistard P., Barloy-Hublier F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Capela D., Becker A., Boutry M., Cadieu B., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kise E., Lelaurie V., Maury D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rampeger U.,
RA Renard C., Thebaud F., Vandenbol W., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RT Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
RN [2]
RP SEQUENCE OF 1483 FROM N.A.
RC STRAIN=1021;
RA Bent A.F., Signer E.R.;
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 1 guanylate cyclase domain.
CC -1- SIMILARITY: TO M.LEPPAE MLCB2407.09.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO
CC FRAMESHIFTS IN POSITIONS 311 AND 321.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL; AL591789; CAC46674.1;
DR EMBL; M30934; AA88525.1; ALT FRAME.
DR InterPro; IPR001054; G_cyclase.
DR SMART; SM00044; CYC; 1.
DR PROSITE; PS00125; GUANYLATE CYCLASES 2; 1.
KM Hypothetical protein; ATP-binding; complete proteome.
FT DOMAIN 39 168 GUANYLATE CYCLASE.
FT NP_BIND 261 268 ATP (POTENTIAL).
FT CONFLICT 323 323 A -> P (IN REF. 2).
SQ SEQUENCE 1029 AA; 112709 MW; 4E4478188B601958 CRC64;

```

```

Query Match 2.8%; Score 110.5; DB 1; Length 1029;
Best Local Similarity 19.0%; Pred. No. 4.8;
Matches 169; Conservative 98; Mismatches 279; Indels 343; Gaps 42;

Cy 17 GAGGAAGDEARAAVRLVRLGPG-----AADSVERLAAPGD 63
Db 26 GGSSTVGGERR---IVTALCYDLVSTDLNVMYDIEDYQELMSAFQLSKQALASHSGVM 82
Cy 64 TSLGGGGA---RVKVGSTGVAAGALHRYLDFC----- 97
Db 83 QHEADGGVALPPIELKKAASLAIRAGLG---IVEACKVRVREGQDDLQVRVGIATSV 140
Cy 98 ---GCHYAM---SGQLRPPRLPAV---PGLTATNRYRYQNVCTQSSPFWM 145
Db 141 ALVLESGRGGWREPVTAAALMAARLQATPNSVLSETRH---LGRSTAFVF- 194
Cy 146 DWARWEREIDWALNGINLALAMSGQEAIWQVYALGLTQAEINEPFGPAFLWGMG 205
Db 195 -----QGSKEKGFAPBEKVMRALGHKVGVD-----FYAFGRMG 229
Cy 206 -----NLHT---WDGPLPSPWHIKQLYIQ-----HRVLDQNR----- 234
Db 230 GPLINRENEINLTIGQLWDGVLAGQGSV---VLIGDAGIGKSRLLRIRTRAKSKLLF 287
Cy 235 -----SFGMTPLYAPAGHVPFAVTRVPQVAVTKKSGWGHFNCSYSCSFLIAPEDPIPI 290
Db 288 PCLUGGGRSTHPLANLPERAVS-----GSAGQMG-----PTAA 323
Cy 291 IGSFLRELKKEFGT---DHLYGADTFENQPPSSPSYLAATTAVYEMTAVTTEAV 346
Db 324 VALTEFRNGIRDPADVDFSYLGAQ--SRLOQSNEDPXAIEKARALLALEAVCR- 380
Cy 347 WLQGMQLGHQPGFMPGPAQI-----RAVLGAVPR--GRLVLDLFAESQOPY- 391
Db 381 -----GPAVAVEDVHMIDPISRDLSGEAALIAKFPVL-LYTSRSBYA 424
Cy 392 -----TRTASFOGQPIWCMHNFGNGLFGALBAVNGPE-- 428
Db 425 SEMLDAAPTRLALPLDSDERLAIKAWP-----ERLALLPLDPATERISGVLEF 479
Cy 429 -----AARLPNSTWVGCMABEGISQNE-VVYSMAELGMRKD-----P 467
Db 480 IEELQWVSQVPEPTMLSEBANSVSAFESILSRLOQIGTRERVAARAAVAGTQVT 539
Cy 468 VPDLAAMV-----TSFAAR-----RYGVSHPDAGAAARLLRSYVN 503
Db 540 LPLRALLPDPEKSLANAAADTLCTGFLTRIRVGRITAGPRH-----TLIGETLYN 592
Cy 504 CSGEACRGHNRSPVLRPSLQNTSITWNSDVPEAWFLILTSAPSLATSPAFRYDLIDL 563
Db 593 ---AVLRKQRYVLRHRR---LFTAVNQNRG--MAAM-----IDT 622
Cy 564 TQAVQELVSLYEARSAYLSKELASLIPAGVLAAYLLPALDEVLASDGRFLGSMLE 623
Db 623 GALA-----EHAERAGLVEAVPLFIAG-----KESSRSAMIERQLE 663
Cy 624 QAPAAVSEADPFYQNSRYQLTLWGPEGNILDYANKOLAGLVANYTTPMRRLFLEALV 683
Db 664 HALDLCQGSSEDTAEALKLALTLGPF-----ILIGVGLSEEPARRLYEDAV- 712
Cy 684 DVAQGIPI-QGHQF-----DKNYFO--LEQAFVLSQKQRP 716
Db 713 -DIARRRPSQSQWPIYWGWLITGQDFRVMDRALVRSLSKANE 760

RESULT 6
ID RHSA_ECOLI STANDARD; PRT; 1377 AA.
AC P16916;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RhsA protein precursor.

```

GN RHSA OR B3593.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxId=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12.  
 RX MEDLINE=90094253; PubMed=2403547;  
 RA Feulner G., Gray J.A., Kirschman J.A., Lehner A.F., Sadosky A.B.,  
 RA Vlazny D.A., Zhang J., Zhao S., Hill C.W.;  
 RT "Structure of the rhaA locus from Escherichia coli K-12 and  
 RT comparison of rhaA with other members of the rha multigene family.";  
 RL J. Bacteriol. 172:446-456(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=94316500; PubMed=8041620;  
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
 RT region from 76.0 to 81.5 minutes.";  
 RL Nucleic Acids Res. 22:2576-2586(1994).  
 RN [3]  
 RP REVIEW.  
 RX MEDLINE=95020608; PubMed=7934896;  
 RA Hill C.W., Sandt C.H., Vlazny D.A.;  
 RT "rha elements of Escherichia coli: a family of genetic composites  
 RT each encoding a large mosaic protein.";  
 RL Mol. Microbiol. 12:865-871(1994).  
 CC -1- FUNCTION: RHA ELEMENTS HAVE A NONESSENTIAL FUNCTION. THEY MAY  
 CC PLAY AN IMPORTANT ROLE IN THE NATURAL ECOLOGY OF THE CELL.  
 CC -1- DOMAIN: EACH RHA APPEARS TO CONSIST OF A HIGHLY CONSERVED 141 kDa  
 CC AMINO FRAGMENT FOLLOWED BY A HIGHLY DIVERGENT CARBOXY TERMINUS.  
 CC -1- SIMILARITY: BELONGS TO THE RHA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; L19044; AAC95065.1; -;  
 DR EMBL; U00039; AAB18570.1; -;  
 DR EMBL; AE000437; AAC76617.1; -;  
 DR PIR; C65159; C65159.  
 DR EcoGene; EGI0846; rhaA.  
 DR InterPro; IPR001826; RHA.  
 DR InterPro; IPR006530; YD.  
 DR Pfam; PF03527; RHA; 1.  
 DR PRINTS; PR00394; RHA; PROTEIN.  
 DR TRIPFAM; TIGR01643; YD repeat 2x; 15.  
 KW Signal; Multigene family; Transmembrane; Repeat; Complete proteome.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1377  
 FT TRANSMEM 28 55  
 FT DOMAIN 330 1186 28 X APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 330 352 1.  
 FT REPEAT 353 374 2.  
 FT REPEAT 375 417 3.  
 FT REPEAT 418 438 4.  
 FT REPEAT 439 460 5.  
 FT REPEAT 461 481 6.  
 FT REPEAT 482 502 7.  
 FT REPEAT 503 525 8.  
 FT REPEAT 526 546 9.  
 FT REPEAT 547 567 10.  
 FT REPEAT 568 588 11.  
 FT REPEAT 589 609 12.  
 FT REPEAT 610 629 13.  
 FT REPEAT 630 650 14.  
 FT REPEAT 651 671 15.

FT REPEAT 672 691 16.  
 FT REPEAT 692 711 17.  
 FT REPEAT 712 734 18.  
 FT REPEAT 735 758 19.  
 FT REPEAT 808 828 20.  
 FT REPEAT 829 850 21.  
 FT REPEAT 851 871 22.  
 FT REPEAT 872 894 23.  
 FT REPEAT 895 930 24.  
 FT REPEAT 931 959 25.  
 FT REPEAT 960 984 26.  
 FT REPEAT 985 1019 27.  
 FT REPEAT 1162 1186 28.  
 SQ SEQUENCE 1377 AA; 156320 MW; 21ACA989E74200FE CRC64;  
 Query Match 2.8%; Score 109.5; DB 1; Length 1377;  
 Best Local Similarity 20.8%; Pred. No. 8.4;  
 Matches 128; Conservative 58; Mismatches 209; Indels 219; Gaps 35;  
 QY 57 AAKPGDITSLGGG---GAARVVRSGTGYAA---AGL--HRYLDPGCHVAMSGSL 108  
 DB 6 AARQG-DMTQYGSIVQSGAGVRIIGAPTVACSVCGVTSQHPVPLLGAKYLPGETDI 64  
 QY 109 RLPRPLPAV-----PG-----ELTEAT-----PNEYRYQ 133  
 DB 65 ALPGPLFILSRYSYRTTPAPVSLGPGWKMGPADILQLRDNLILSDNGRELVE 124  
 QY 134 NVC--TQSYFVWDMARWEREIDMANGINIALAMSGOEAIQSV-----YLA 181  
 DB 125 HLPFGEDGVS-----RSESLW-VRGGAALDEGHRLLAALMQALPELRLSPRYIA 175  
 QY 182 LGTQAEINFFTPGPAFLAMGRMNLHTW-----DGLPPSMWIKQYLQHRVL 230  
 DB 176 TNSPQ-----GP-----MWLIG--MCEVPEADDEVLPALPP-----YRVL 209  
 QY 231 DQW-RSFGMTPLVPAPAGHVEAVTVFPQVNTKMGSGHFNCSYSCFLAPEDPIFP 289  
 DB 210 TGVDFRGRQTQFHR-----EAGFSGEITGVTDGAWHFRLVLTQORALE----- 258  
 QY 290 IIGSLFRELKEKFGDHTYGAADTFNEMQPSSEPEYLAATVAEAMTAUVEAWTL 349  
 DB 259 -----ARQALSGGTE-----PSAFPTLLPGYTE--YGRDNGILSLAWL- 296  
 QY 350 QGWLFOHQPQFMGPAPIRAV---LGAVPRGLVLIDLPASQPVYTRTASFGQGFPIWC 405  
 DB 297 -----THDPEY--PENLPAPPIVRYGMPTRGELAV-----VYDRSGK- 311  
 QY 406 MANNFGGNHGLFALDA--VNGCEPARLFPNSITVGTGMAPEGIS----- 449  
 DB 332 QVRSFTYDDKYRGMVAHRTGTGFEIRYRYSQDGTQOLNPAGLSYTYQEKORITTD 391  
 QY 450 ---ONEVVSIMAEIGMRK-----DVPDLAMVTSFAARRYGVSHPDA 490  
 DB 392 SLRREVLRH-QGEAGLKVYKKEHADSGVTQSGFPAVGRIR--QTDAAGRITTESPDV 448  
 QY 491 GAARLLILSVNCSGECAC--RGHNSPLVRRP-SLOMNTSITWNRSDVPEAMFLILT 545  
 DB 449 VTG---LITFITTPDRASAFYVNHNOULTSATGPDGLEI-----REYDELGLIOE 498  
 QY 546 SAPSLATSPAFRYD 559  
 DB 499 TAPD-GDITRYRYD 511  
 RESULT 7  
 ID RHB\_ECOLI STANDARD; PRT; 1411 AA.  
 AC P16917; P76701;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE RHB protein precursor.  
 GN RHB OR B3482.

OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OC NCBI\_TaxId=562;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RC MEDLINE=93259920; PubMed=8387990;  
 RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;  
 RT "Rhs elements of Escherichia coli K-12: complex composites of shared  
 RT and unique components that have different evolutionary histories.";  
 RI J. Bacteriol. 175:2799-2808(1993).  
 RN [2]  
 RN REVISION TO 405.  
 RA Hill C.W.;  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
 RP [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RC MEDLINE=94316500; PubMed=8041620;  
 RA Sofia H.U., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
 RT region from 76.0 to 81.5 minutes.";  
 RL Nucleic Acids Res. 22:2576-2586(1994).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RC MEDLINE=9742617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [5]  
 RP SEQUENCE OF 1-100 FROM N.A.  
 RC STRAIN=K12;  
 RC MEDLINE=89123133; PubMed=2644231;  
 RA Sadosky A.B., Davidson A., Lin R.J., Hill C.W.;  
 RT "rhs gene family of Escherichia coli K-12.";  
 RL J. Bacteriol. 171:636-642(1989).  
 RN [6]  
 RP SEQUENCE OF 1221-1411 FROM N.A.  
 RC STRAIN=K12;  
 RC MEDLINE=90094253; PubMed=2403547;  
 RA Feulner G., Gray J.A., Kirschner J.A., Lehner A.F., Sadosky A.B.,  
 RA Vlazny D.A., Zhang J., Zhao S., Hill C.W.;  
 RT "Structure of the rhsA locus from Escherichia coli K-12 and  
 RT comparison of rhsA with other members of the rhs multigene family.";  
 RL J. Bacteriol. 172:446-456(1990).  
 RN [7]  
 RP REVIEW.  
 RP MEDLINE=95020608; PubMed=7934896;  
 RA Hill C.W., Sandt C.H., Vlazny D.A.;  
 RT "Rhs elements of Escherichia coli: a family of genetic composites  
 RT each encoding a large mosaic protein.";  
 RL Mol. Microbiol. 12:865-871(1994).  
 CC -1- FUNCTION: RHS ELEMENTS HAVE A NONESSENTIAL FUNCTION. THEY MAY  
 CC PLAY AN IMPORTANT ROLE IN THE NATURAL ECOLOGY OF THE CELL.  
 CC -1- DOMAIN: EACH RHS APPEARS TO CONSIST OF A HIGHLY CONSERVED 141 KDA  
 CC AMINO FRAGMENT FOLLOWED BY A HIGHLY DIVERGENT CARBOXY TERMINUS.  
 CC -1- SIMILARITY: BELONGS TO THE RHS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L02370; AAC61883.1; -  
 DR EMBL; U00039; AAB18457.1; -

DR EMBL; AB000424; AAC76507.1; -  
 DR PIR; E65145; E65145.  
 DR Ecogene; EG10847; rhsB.  
 DR InterPro; IPR001826; rhsB.  
 DR InterPro; IPR006530; YD.  
 DR Pfam; PF03527; rhsB; 1.  
 DR PRINTS; PR00394; rhsBPROTEIN.  
 DR TIGRfam; TIGR01643; YD\_repeat\_2x; 15.  
 KW Signal; Multigene family; Transmembrane; Repeat; Complete proteome.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1411  
 FT TRANSMEM 28 55  
 FT DOMAIN 330 1186  
 FT REPEAT 330 352  
 FT REPEAT 353 374  
 FT REPEAT 375 417  
 FT REPEAT 418 438  
 FT REPEAT 439 460  
 FT REPEAT 461 481  
 FT REPEAT 482 502  
 FT REPEAT 503 525  
 FT REPEAT 526 546  
 FT REPEAT 547 567  
 FT REPEAT 568 588  
 FT REPEAT 589 609  
 FT REPEAT 610 629  
 FT REPEAT 630 650  
 FT REPEAT 651 671  
 FT REPEAT 672 691  
 FT REPEAT 692 711  
 FT REPEAT 712 734  
 FT REPEAT 735 758  
 FT REPEAT 759 784  
 FT REPEAT 808 828  
 FT REPEAT 829 850  
 FT REPEAT 851 871  
 FT REPEAT 872 894  
 FT REPEAT 895 930  
 FT REPEAT 931 959  
 FT REPEAT 960 984  
 FT REPEAT 985 1019  
 FT REPEAT 1162 1186  
 FT CONFLICT 1130 1130 Q -> K (IN REF. 4).  
 SQ SEQUENCE 1411 AA; 159394 MM; 567AC4EB713D9E07 CRC64;  
 Query Match 2.8%; Score 109.5; DB 1; Length 1411;  
 Best Local Similarity 20.8%; Pred. No. 8.6;  
 Matches 128; Conservative 58; Mismatches 209; Indels 219; Gaps 35;  
 QY 57 AKKPGLDIYSLGG---GAARVVRGSGVAA---AGL--HRYLRDSCGHVAMSGSL 108  
 DB 6 AARQG-DMTQYGSIVQSGAGVRIGAPFGVACVCPGAVTSGHPVNPILGAKVLPGETDI 64  
 QY 109 RLPRPLPAV---PG-----ELTEAT-----PNEYRYQ 133  
 DB 65 ALPGPLPFLISTSYRTKTPAPVGLSPGKQKPADIRLQLRDNTLILSDNGSLTYFE 124  
 QY 134 NYC--TQSYSFVPMWMAKMERIDMALNGINLALAMSGEALNQRV-----YLA 181  
 DB 125 HLPFSDGYS-----RSSSLVLVGVGAKLDEGRLAALWQALPEELRSLSPHYLA 175  
 QY 182 LGLTQALINEPFTGPAFLAMGMNLHTM-----DGLPSPWHIKOLYLQHRVL 230  
 DB 176 TNSPQ-----GF-----WMLLG-----WCERVPEADEVLPAFLP 209  
 QY 231 DQM-BSFGKTPVLPAPAGHVPATVTRVPQVNTKMGSGWGFNCYSGSFLAPDPFP 289  
 DB 210 TGLVVRFGRTQTFRH-----EAAEFGSGEITGVTDGMRHRLVLTQQAQPAE 258  
 QY 290 IIGSLFLRELKEFGDHIYGADTFENMQPSSSESYLAATTAVEAMTAVTEAVWL 349  
 DB 259 -----ARQQAISGTE-----PSAFPDTLPQTE--YGDNGIRLSAVWL- 296  
 QY 350 QGMLFQHPQFVWGAQIRAV---LGAVPGRRLVLDLFAESQPFYTRTASFQGPFTWC 405

```

Db 297 ----THDPFY--PENLPAALPVYXWTFRGELV-----VYDRSGK----- 331
Qy 406 MHNFGNGHGLFGALFA--VNGGPEAARLFNFTWGTGNAPGIS----- 449
Db 332 QVRFSTYDDKRGKGMVNRHRTGRPRIRYRSDGRVTEQLNPAALSTTYQEKDRITTD 331
Qy 450 ---QNEVYYSIMAEIGWRK-----DVPDLAAWVTSFAARYGVSHDA 490
Db 392 SLDRREVLTHT-QGEAGLGRVYKKEHADGVSQSFDAVGRIFA--CTDAAGRTTESPDV 448
Qy 491 GAAMRLIRSVYNSGEGAC-----RGHNSPLVRRP-SLQWMTSTWNRSDVFEAMRLILT 545
Db 449 VTG--LITRTITPDGRASAFYVNHNDLTSATGPDGLE-----RREIDELGRLIOE 498
Qy 546 SAPSLATSPAFRYD 559
Db 499 TAPD-GDITRRYD 511

RESULT 8
CATA_SALTY STANDARD; PRT; 726 AA.
ID CATA_SALTY
AC P17750;
DT 01-AUG-1990 (Rel. 15, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Peroxidase/catalase HPI (EC 1.11.1.6) (Catalase-peroxidase)
DE (Hydroperoxidase I).
DE KATG OR STM4106.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=91117169; PubMed=2277629;
RA Loewen P.C., Stauffer G.V.;
RT "Nucleotide sequence of katG of Salmonella typhimurium LT2 and
RT characterization of its product, hydroperoxidase I."
RL Mol. Genet. 224:147-151 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSCL412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Potwolk S., Ali U., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RN Nature 413:852-856 (2001).
RC -1- FUNCTION: Bifunctional, exhibiting both a catalase and
RC broad-spectrum peroxidase activities.
CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC -1- COFACTOR: Binds 2 prohememe IX and 2 iron ions per tetramer.
CC -1- SUBUNIT: Homotrimer.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
CC -1- SIMILARITY: Contains 1 heme regulatory motif (HRM) repeat.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X53001; CAA37187.1; -
DR EMBL; AE008891; AAL22946.1; -

```

```

DR PIR; S12039; CSEBHT.
DR HSSP; P48534; IAPX.
DR StryGene; SG10190; KatG.
DR InterPro; IPR000763; Bac. class/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase; 1.
DR PRINTS; PRO0486; PEROXIDASE.
DR TIGRfam; TIGR00198; cat_per_HPI; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00436; PEROXIDASE_2; 1.
DR PROSITE; PS00873; PEROXIDASE_4; 1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;
KW complete proteome.
KM Complete proteome.
FT ACT_SITE 102 102
FT ACT_SITE 106 106
FT ACT_SITE 267 267
FT METAL 267 267
FT REPEAT 15 20
FT CONFLICT 71 71
FT CONFLICT 136 136
FT CONFLICT 223 223
FT CONFLICT 228 228
FT CONFLICT 233 233
FT CONFLICT 233 233
FT CONFLICT 275 275
FT CONFLICT 545 545
FT CONFLICT 549 549
FT CONFLICT 553 553
FT CONFLICT 561 561
FT CONFLICT 634 634
SQ SEQUENCE 726 AA; 79656 MW; 7C4BA4439E9FAB CRC64;

Query Match 2.7%; Score 108; DB 1; Length 726;
Best local similarity 17.7%; Pred. No. 4.6;
Matches 134; Conservative 73; Mismatches 196; Indels 352; Gaps 33;

Qy 86 AAGLHYRLDFCGCHVANSQQLRPLPAVPGELTEATNR-----YRYNQVCTQ 138
Db 26 SAGAGTASRD-----WPNQLRV-----DLNQHNSRNPGLGEDPYRKEFKL 69
Qy 139 SYSFV-----WMDARERELDMNALINGIMLAMS----- 169
Db 70 DYSALKGDIKALLTTSQPMW-----PADMGSYGLTIRNANHAGYRISDGRGAAR 122
Qy 170 -----GQEAIWQVYIALGLTQAEINEFFTPAF 198
Db 123 GQORFAPLNSWPDNVLDPKARLLPWIKOKYQOKISMDLFLIAGNVALENGSEFT---- 178
Qy 199 LAMGGMNLHTWDGLPSPWHIKOYLQHRVLDQMRSGMTFVLPFAFGHVEATRVFP 258
Db 179 PGGG-AGREDVPEPLDVWGDGKALTHR-----H-PEALAKA-- 215
Qy 259 QVAVTKMGSGWGHFNCSYSCSFLAPEDPIFLIGSLFREDIKERTDHIYGADTFNEMQ 318
Db 216 PLGATEMG-----LIYVNEGP-----DH----- 234
Qy 319 PRSSSPSYLAATTVAYEAMTVADTEAVWLLQGMFLPQHPQEWGPAQIRAVLGAIVPRGL 378
Db 235 --SGEPLAALAAIRATFGMGKNDDETVALLAG--GHTLKGTHGAALASHVADP-- 285
Qy 379 LVLDLFAASQPYVYRTASTAFQGGPFIWCMLNHFGNGHGLFGALVANGGPEAA-- 430
Db 286 -----EAPF-----EAQGLGMASYSQG--VGA-DATSLSEVWVWTPQTFFQ 325
Qy 431 -----RLFPNSTWGTGMAPEGISQNEVYS-----LMAELGMKRD 466
Db 326 SNYFFENLFKXEMV--QTRSPGALQFAVAVAPDILPPFPSPKRRKPTMTVTLTLFDF 383
Qy 467 P-----VFDLAAMVTSFAARYGVSHPDAGAAWMLLRSVYNSGEGACRGNRSPL 517
Db 384 PEFKTSRRFLNDPQAFNPAFAKFKLTHDMGKAYI----- 423
Qy 518 VARPSLQWNTSIWNRSDVFEAMRLILTSAPSLATSPAFRYDLDL-----TROAVOEL 571
Db 424 --GPVPEKEDLWQD-----PLPQPLTYQPT-QEDIIWLKAAIAGSGISSEM 467

```

```

QY 572 VSLYVEEARSAYLSKELASLLRAG-----GVLAYELLPAIDVYL 610
Db 468 VSV-----AMAS---ASTFRGDDKRGANGARLALPORDVNAVAARVLPVLEKQ 517
QY 611 ASDSRFLGDS-----WLEQARAAA-----VSEAEAD 636
Db 518 KTKVKAHLADIIVLAGVGIQAAAAGVSIYFAPRGVDAQDQDIEMFSLPIAD 577
QY 637 FYED-----NSRYOLTMGEGNUL 656
Db 578 GFENVYRRLDVSTTESLILDKAQQLTLPAPMTVL 612

RESULT 9
CATA_ECOLI STANDARD; PRT; 726 AA.
AC P13029; 01-JAN-1990 (Rel. 13, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Peroxidase/catalase HPI (EC 1.11.1.6) (Catalase-peroxidase)
DE (Hydroperoxidase 1).
GN KATG OR B3942.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=562;
[1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=88314956; PubMed=3045098;
RA Triggs-Raine B.U., Doble B.W., Mulvey M.R., Scoby P.A., Loewen P.C.;
RT "Nucleotide sequence of katG, encoding catalase HPI of Escherichia
RT coli."
RL J. Bacteriol. 170:4415-4419 (1988).
[2]
RP SEQUENCE OF 1-339 FROM N.A.
RX STRAIN-K12 / MG1655.
RX MEDLINE=93347969; PubMed=8346018;
RA Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R.;
RT "Analysis of the Escherichia coli genome. III. DNA sequence of the
RT region from 87.2 to 89.2 minutes."
RL Nucleic Acids Res. 21:3391-3398 (1993).
[3]
RP SEQUENCE OF 309-726 FROM N.A.
RX STRAIN-K12 / MG1655.
RX MEDLINE=94089392; PubMed=8265357;
RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA Daniels D.L.;
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT region from 89.2 to 92.8 minutes."
RL Nucleic Acids Res. 21:5408-5417 (1993).
CC - FUNCTION: Bifunctional, exhibiting both a catalase and
CC broad-spectrum peroxidase activities.
CC - CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC - COFACTOR: Binds 2 prothemo IX and 2 iron ions per tetramer.
CC - SUBUNIT: Homotetramer.
CC - INDUCTION: By hydrogen peroxide.
CC - PMT: The N-terminus is blocked.
CC - SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
CC - SIMILARITY: Contains 1 heme regulatory motif (HRM) repeat.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M21516; AAA24040.1; -
DR EMBL; L19201; AAB03074.1; -

```

```

DR EMBL; U00006; AAC43048.1; -
DR EMBL; AE000468; AAC76924.1; -
DR PIR; A65201; CSECHP.
DR HSSP; P00431; 1CYF.
DR SWISS-2DPAGE; P13029; COLI.
DR Ecogene; EG10511; KatG.
DR InterPro; IPR000763; Bac_ctase/prase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00141; peroxidase_1.
DR PRINTS; PRO0458; PEROXIDASE.
DR TIGRfam; TIGR00198; cat_per_HPI_1.
DR PROSITE; PS00435; PEROXIDASE_1.
DR PROSITE; PS00436; PEROXIDASE_2.
DR PROSITE; PS00873; PEROXIDASE_4.
DR Oxidoreductase; peroxidase; Iron; Heme; Hydrogen peroxide;
KW Complete proteome.
FT ACT SITE 102 102 BY SIMILARITY.
FT ACT SITE 106 106 BY SIMILARITY.
FT METIL 267 267 IRON (HEME AXIAL LIGAND).
FT REPEAT 15 20 HRM.
FT CONFLICT 621 621 A -> G (IN REF. 1).
SQ SEQUENCE 726 AA; 80023 MW; 24D32EBED5D59BD6 CRC64;

Query Match 2.7%; Score 104.5; DB 1; Length 726;
Best Local Similarity 18.3%; Pred. No. 8.2;
Matches 125; Conservative 66; Mismatches 169; Indels 303; Gaps 32;

QY 86 AAGLHRVYRDPGCGHVAAGSGQLRPLPRAVPGELTEATNR-----YR----- 130
Db 26 SAGAGTTRD-----WMPNQLRV-----DLNQHNRNPNJGEDPDKERSKL 69
QY 131 -YY-----QNVCTOSYFVWMDARWEREIDMNLNGINLALMS----- 169
Db 70 DYGLKAKLKLITRSQ--W-----PADMGSYAGLFRMAHNGAGTYSIDRGGA 120
QY 170 -----GQAIWQRYTALGLTQAINETFTGP 196
Db 121 GRCQRFAPLNSWPDNVLDRARLLWPIKQYGGKISWALFLIAGVALENGFRV- 178
QY 197 AFLANGRGNIHTMDGRLPSPMHIKQLYLQHRVLDQNRFSQTPVLPFAAGHVAEAVRV 256
Db 179 -FGFG-AGREDVWEPDLDVWMDKAVLTR-----H-PALAKA 215
QY 257 PQVAVVTWGGSWGHCNCSYCSFLAPDPPIPTIGSLFLRELKEFGTDHIYGADTFNE 316
Db 216 -PLGATENG-----LIYVPEGP-----DH----- 234
QY 317 NQPSSESYLAATTAYEAMTAVDTAAVWLQGWLFQHPQFQPAQIAVIGAVPRG 376
Db 235 ---GGEPLSAAALIRATFGNMGNDDETVALLIAGHTLGKTHGAGPT--SNVGPDP-- 285
QY 377 RLIVLDLFAESQPVYTR-----TASFOGQPIWCMILHNGGNH--GLF-- 417
Db 286 -----EAPRIEGLQMGASTYSGVGADATISGLEVWVTQPTQMSVYFENLPKY 336
QY 418 -----GAL--EAVNGGPEAARLFPNSTWVGCMAPBEGISQNEVYYSIMAEIGWKD 466
Db 337 EMVQTRSPAGALQFRAV-----DAPEIIPD-----PFDPSKKRKFTMTVTDLTLEFD 383
QY 467 P-----VPLDAWTSFAARRYGVSHPDAGAWMLLRSYVNCSGEACRGNRSP 517
Db 384 DEFELKSRFLNDPQAFNEAPARAFKLTTHDMGKSYI----- 423
QY 518 VERPQLQWNTSIWNRSDVFEARLLTTSAPSLATSPAFRDLDDL-----TROAVQEL 571
Db 424 -GPEVPEKDLTWQD-----PLQPIYVPL-EDDIIDLKFAIDSGUSVEL 467
QY 572 VSLYVEEARSAYLSKELASLLRAG-----GVLAYELLPAIDVYL 610
Db 468 VSV-----AMAS---ASTFRGDDKRGANGARLALPORDVNAVAARVLPVLEKQ 517
QY 611 ASDSRFLGDSWLEQARAAA-----VSEAEAD 636

```

DB 518 KESKASLADITVAGVGEKA 540

RESULT 10

P85B RAT STANDARD; PRT; 722 AA.

AC 063788;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Phosphatidylinositol 3-kinase regulatory beta subunit (PI3-kinase p85-beta subunit) (Ptdins-3-kinase p85-beta).

DE PI3K2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RP SEQUENCE FROM N.A.

RC STRAIN=Wistar; TISSUE=Brain;

RX MEDLINE=96214979; PubMed=8621382;

RA Inukai K., Anai M., Vandreva E., Hosaka T., Katagiri H., Funaki M., Fukushima Y., Ogihara T., Yazaki Y., Kikuchi M., Oka Y., Asano T.;

RT "A novel 55-kDa regulatory subunit for phosphatidylinositol 3-kinase structurally similar to p55PIK is generated by alternative splicing of the p85alpha gene.";

RL J. Biol. Chem. 271:5317-5320(1996).

CC - FUNCTION: BINDS TO ACTIVATED (PHOSPHORYLATED) PROTEIN-TYROSINE KINASES, THROUGH ITS SH2 DOMAIN, AND ACTS AS AN ADAPTER, MEDIATING THE ASSOCIATION OF THE P110 CATALYTIC UNIT TO THE PLASMA MEMBRANE.

CC - SUBUNIT: HETERODIMER OF A P110 (CATALYTIC) AND A P85 (REGULATORY) SUBUNIT.

CC - SIMILARITY: BELONGS TO THE PI3K P85 SUBUNIT FAMILY.

CC - SIMILARITY: Contains 1 Rho-GAP domain.

CC - SIMILARITY: Contains 2 SH2 domains.

CC - SIMILARITY: Contains 1 SH3 domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.jsb-sib.ch/announce/> or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).

CC -----

DR EMBL; D64046; BAA10926.1; -

DR HSSP; P23727; 2PNB.

DR InterPro; IPR001720; PI3K\_kinase\_P85.

DR InterPro; IPR000198; RhoGAP.

DR InterPro; IPR000980; SH2.

DR InterPro; IPR001452; SH3.

DR Pfam; PF00620; RhoGAP; 1.

DR Pfam; PF00017; SH2; 2.

DR Pfam; PF00018; SH3; 1.

DR PRINTS; PR00678; PI3KINASEP85.

DR PRINTS; PR00401; SH2DOMAIN.

DR PRODOM; PD000093; SH2; 2.

DR SMART; SM00324; RhoGAP; 1.

DR SMART; SM00252; SH2; 2.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS50238; RHO-GAP; 1.

DR PROSITE; PS50001; SH2; 2.

DR PROSITE; PS50002; SH3; 1.

DR SH3 domain; SH2 domain; Repeat.

FT DOMAIN 4 80 SH3.

FT DOMAIN 112 289 RHO-GAP.

FT DOMAIN 324 419 SH2.1.

FT DOMAIN 616 710 SH2.2.

SO SEQUENCE 722 AA; 81328 MW; 12083688B9F670C55 CRC64;

Query Match 2.6%; Score 103; DB 1; Length 722;

Best Local Similarity 21.6%; Pred. No. 10;

Matches 87; Conservative 38; Mismatches 134; Indels 144; Gaps 19;

QY 74 RVRVNGSTGVAAAAGLHRYLDFCGGV-----AMSGQLRPLRLPAVF----- 118

DB 63 RTRRG-----DFPGTYVEFLGPAVALAPGPRPRGRPLPAPFLDGPSS 107

QY 119 -----GELTE-----ATPNRRYRYQNVCTOSYSFVW-----DMA-----RWER 152

DB 108 GHTLASLAEQSPSPSAPPIVLKLEAIQALDESEFSPRLPAPRTDMSLDEQMDR 167

QY 153 EIDWMAINGINIALAMSGQEAIWQRYVALGITQAEINEFF-----TGPAFLANGRMGN 206

DB 168 TLLYDAVKGFLALP-----AAVTPEAAAEAYANAEVIGPVGLVLE----- 210

QY 207 LHTWDGPIPLPSWHIKQVLYQHRYVLDQMSFGMTPLPAPAGHPVAVRVFQVAVTKMG 266

DB 211 -----PTPLPLQALTLRPLQLHGVARRAPSPATVH-----ALASAFPL----- 253

QY 267 SWGHFNCSYSCSFLAPED-----PIFPIIGSLFRLRLKEFGTDHIYGDTPNE 316

DB 254 -----LLNAPPGEGDSEPADEPV-----LLERLVGE-----HVDEQDTAP 294

QY 317 MPPSPSESYLAATTAYIE-AMTAVDTEAVVLLQGLFQHPQVQWP---AQIAYVGA 372

DB 295 ALPP--KSKVAPAPETALANGSTSLQDAEW-----YMGDISREEVNERLUD 340

QY 373 VPRGRLLVLDLPAESQPVYTRTASFGQPFICMLNFGNNG 415

DB 341 TDDGTFLVRDASKIQGEYTLTRKGNKKLIKVRH-DGHYG 382

RESULT 11

YMCA ECOLI STANDARD; PRT; 698 AA.

ID YMCA ECOLI

AC P75882;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical 11proprotein ymca precursor.

GN YMCA OR B0984.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

RA "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RT [2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=97061202; PubMed=8905232;

RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kaishimoto K., Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizubuchi K., Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N., Sampaio G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano W., Horuchi T.;

RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map.";

RL DNA Res. 3:117-155(1996).

CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor (Probable).

CC -1- SIMILARITY: STRONG, TO E. COLI YUBH.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AB000200; AAC74069.1; -  
 DR EMBL: D90735; BAA35749.1; -  
 DR EMBL: D90736; BAA36124.1; -  
 DR PIR: F64839; F64839.  
 DR EcoGene; EG13729; ymcA.  
 DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.  
 KM Hypothetical protein; Membrane; Lipoprotein; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 698 HYPOTHETICAL LIPOPROTEIN YMC.  
 FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL).  
 SQ SEQUENCE 698 AA; 78687 MW; 37F71D2DC57FA4E7 CRC64;

Query March 2.6%; Score 101; DB 1; Length 698;  
 Best local similarity 19.2%; Pred. No. 14; Index 300; Gaps 39;  
 Matches 146; Conservative 75; Mismatches 239;

QY 108 LRLRPPAVVGE--LLEATPNRYRYQ-----NYCTOSYSPVWMDMA- 148  
 DB 43 LQWNAIAPGEGSVNYRNDQYRFSYSTVALFPMLEGTIRYTDVTRKYS-QMEDFSG 101  
 QY 149 -----RMRERIDMAINLAINLAMSQGEALWCQVYIALGL-----T 185  
 DB 102 DQSYKDSFDFKRLMEEGY-----WLPVAGKKGDIAGTGLFDGEYLVASK 148  
 QY 186 QAEINEFFTPGPAFLAMRGKMLHTWDGRLPSPWHIKOLYLOHNV-----LDQMS 235  
 DB 149 QAGSPDTLIG--MAMGYAGNAGNITMP--CRVSDKVC-HAASHDAGDISFSDIFFG 201  
 QY 236 ---FG--MTPLVLA-----FAGHVEAVTRVFPQVNTTKMGSGHENC 273  
 DB 202 PASIFGIEYQTPMNPRLKLEYDGNVYONDFAKGLPQA--SHFNQAVYRAASMDLNL 259  
 QY 274 SYSOSFLAPEDPFPITIGSLFLRELKERTDHYG--ADTFENMOP--PSSPSYLV 327  
 DB 260 SY-----ERGNILMEFPLTRTNDRPLRTPKPAVQ 293  
 QY 328 AAAT--TAYEANTAV-----DTEAVMLQGMFLQHPQDFWGPACIRAVLGAVR 375  
 DB 294 PAFSEGLQYTTVANQUTALKYNAGDAPAEIQLRDKTLVMSGOQY--KYRDSRAVD 349  
 QY 376 -GRLLVLDLPAESQPVYTRTASFGQGFIMQMLNFGNNGLFGALEAVNGGEARLRF 434  
 DB 350 ANRLVNNL-----P-----QGVETI-----STQKEHMAWT 378  
 QY 435 NSTWVGT-----GMAEGISQNEVVYSLMAELGRKDPVDDLAAMVTSFAARVGVSH 488  
 DB 379 TETIVASIRKQLAGTAP---QOSEPLQOQVVE-----AEDLSAFGRGRINEDPSYS 428  
 QY 489 DAGAAMRLLRSVNCGEACRGHNSPLVRRBPLQNTSIW-----YNR 533  
 DB 429 -----NPFLLSQSLGPEDEPFYFQGLMSARVFTDHLLDGIFETNIYNN 475  
 QY 534 SDVEAWRLILTSAPSLATSPAFYDDLDIRQAV-----QELVSLYTEEARSAV 583  
 DB 476 YDKKSSIL-----PASSTLPKRVTHIRDVYRNDVYLNMQAYFADLNGSFGQVYGV 530  
 QY 584 LSKELASILRAGVLAELLALDEVIASDSRFLLSGWLQAPRAAVALSEADPFYEONR 643  
 DB 531 LETWYAGV--GSELVYRPL-----DACWMLG-----VDVYVYVQRD- 564  
 QY 644 YQLTMGEGNILLVANKQLAGLVANYVTPKRWLFLEALVDVSAQGI PPOHQFDKVPQ 703  
 DB 565 -----WQWMAFTYISFP--TGVTATVWNP-----PILNGVLMK----- 596  
 QY 704 LEQAFVLSKORYPSQPGDITVLAKKIFLKYPGWVAGSW 743  
 DB 597 -----LSVQGYLAKDKGATIDVAK---RDSGVAVAGVW 626

RESULT 12  
 ID 0156 STRAT STANDARD; PRT; 3519 AA.  
 AC 007017;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Oleandomycin polyketide synthase, modules 5 and 6.  
 OS ORF.  
 OS Streptomyces antibioticus.  
 OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycesetaceae; Streptomyces.  
 OK NCBI\_taxid=1890;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94150470; PubMed=8107683;  
 RA Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;  
 RT "Characterisation of a Streptomyces antibioticus gene encoding a type  
 RT I polyketide synthase which has an unusual coding sequence."  
 RL Mol. Gen. Genet. 242:358-362 (1994).  
 CC -1- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN  
 CC LACTONE RING.  
 CC -1- COPOLYMER: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.  
 CC -1- SIMILARITY: Contains 2 acyl carrier domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L09654; AAA16695.1; -  
 DR PIR: S43048; S43048.  
 DR HSSP: P25715; 1MLA.  
 DR InterPro: IPR001227; Ac transferase.  
 DR InterPro: IPR000794; Ketoacyl-synt.  
 DR InterPro: IPR006163; pp-bind.  
 DR InterPro: IPR006162; ppantne\_attach.  
 DR InterPro: IPR001031; Thioesterase.  
 DR Pfam: PF00698; Acyl\_transf. 2.  
 DR Pfam: PF00109; ketoacyl-synt; 2.  
 DR Pfam: PF02801; ketoacyl-synt C; 2.  
 DR Pfam: PF00550; pp-binding; 2.  
 DR Pfam: PF00875; Thioesterase; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.  
 DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 2.  
 DR PROSITE; PS00075; ACP DOMAIN; 2.  
 DR TRANSFERASE; Acyltransferase; Antibiotic biosynthesis; NADP;  
 KW Phosphopantetheine; Multifunctional enzyme; Repeat.

FT DOMAIN 1 3519  
 FT DOMAIN 2 32 501  
 FT DOMAIN 3 569 890  
 FT DOMAIN 4 1200 1382  
 FT DOMAIN 5 1487 1561  
 FT DOMAIN 6 1686 2156  
 FT DOMAIN 7 2220 2541  
 FT DOMAIN 8 2856 3038  
 FT DOMAIN 9 3141 3215  
 FT DOMAIN 10 3270 3519  
 FT ACT\_SITE 210 210  
 FT ACT\_SITE 660 660  
 FT NP\_BIND 1203 1249  
 FT BINDING 1524 1524  
 FT ACT\_SITE 1859 1859  
 FT ACT\_SITE 2311 2311  
 FT ACT\_SITE 2311 2311  
 FT NP\_BIND 2859 2905  
 FT BINDING 3178 3178  
 SQ SEQUENCE 3519 AA; 368561 MW; 41AE78AAAB61F86 CRC64;



CY		319	PPSSPSSTLAATAATVAVYAMNVAVDPEAAWMLQGLFQCPQPWCSPAQRALVAVPRGRLL	378
Db		235	--SSEPLSAANAARTRTFGMGNNDDEYTAALLAG---GHTLGTGTHGAANAASHVGADP-----	285
CY		379	LVLDFPAESQPYVTETASTAQGPFTIWMCHNFNGHGLFGALEAVNGDPEAA-----	430
Db		286	-----AAPF-----EAQGLGMASSYGSG-----VGA-DATISGLAEVMTQTPTQW	325
CY		431	-----RLFPNSTWGCTGAPEGISQNSEVVS-----LMAEIGMRKD	466
Db		326	SNYFFENLFKIEWV--QTRSPGAIQFEAYDAPDIIIPDPFSKRKTMTLVTDLTLPD	383
CY		467	P-----VPDLAWTSPFAARRGYVSHPDGAMWELLRSVNSCGEACRGHNRSPL	517
Db		384	PEFEKISRPFINDPOAFNEAFARAFMKTLTRMGCKRAYI-----	423
CY		518	VRRBELQNTGISIWNRSDVFEMAKRLLLSAPSLSAPFRYDLDTL-----TROAQOEL	571
Db		424	--GEPVPREDILWD-----PLPOLYQPT-QEDITNLKAALNAASGLISEM	467
CY		572	VSLYYEARSAAYLSKEELASLRAG-----GVLAYELLPALDBVL	610
Db		468	VSV-----AMAS---ASTRGDGKKGGANGARALAPQRPDMENVAVAAVRLPYLEALQ	517
CY		611	ASDSRFLLGS-----WLEQARA--A-----VEEABAD	636
Db		518	KTKNKASLADIIVLAGVGIQAANAAAGVISVFPAPGRVDARPDQDTDIEMPSLLEPIAD	577
CY		637	FYEQ-----NSRYQVLTLMGPREGNIL	656
Db		578	GFRNVYRARLDVSTTESLLIDKAQQTLTLAPEMTVL	612
 RESULT 14 ENV_HVZD1 ID ENV HVZD1 STANDARD; PRT; 851 AA.				
AC		P17755;		
DT		01-AUG-1990 (Rel. 15, Created)		
DT		01-NOV-1991 (Rel. 20, Last sequence update)		
DT		15-JUL-1993 (Rel. 38, Last annotation update)		
DE		Envelope polyprotein GP160 precursor [Contains: Exterior membrane glycoprotein (GP120); Transmembrane glycoprotein (GP41)].		
DE		GN ENV.		
OS		Human immunodeficiency virus type 2 (Isolate D194) (HIV-2).		
OC		Vitruves; Retroid vitruves; Retroviridae; Lentivirus.		
CC		NCHI_TaxID=11713;		
RN		[1]		
RP		SEQUENCE FROM N.A.		
RX		MEDLINE=91045094; PubMed=2235509;		
RX		MEDLINE=69184631; PubMed=2467304;		
RA		Kuehnelt H., von Briesen H., Dietrich U., Adamski M., Mix D.,		
RA		Biesert L., Kreutz R., Immelmann A., Henoc K., Melchner C.,		
RA		Andresen R., Geidelblom H., Ruebsamen-Waigmann H.;		
RT		"Molecular cloning of two west African human immunodeficiency virus type 2 isolates that replicate well in macrophages; a Gambian isolate, from a patient with neurologic acquired immunodeficiency syndrome, and a highly divergent Ghanaian isolate."		
RT		Proc. Natl. Acad. Sci. U.S.A. 86:2383-2387(1989).		
CC		-1- MISCELLANEOUS: THIS ISOLATE IS FROM A GAMBIAN CASE OF 'NEURO-AIDS'.		
CC		-----		
CC		THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial		

CC	entities requires a license agreement (See http://www.jsb-sib.ch/announcement)					
CC	or send an email to license@lsb-sib.ch).					
CC	-----					
DR	EMBL; J04542; AAA76847.1; -;					
DR	EMBL; X52223; CAAB6471.1; -;					
DR	PfR; S12159; S12159.					
DR	HIV; J04542; ENV52D194.					
DR	InterPro; IPR000328; Env_GP41.					
DR	InterPro; IPR00777; GP120.					
DR	Pfam; PF00516; GP120; 1.					
KM	AUG; Coac protein; Polypeptidein; Glycoprotein; Transmembrane;					
KW	Signal.					
FT	SIGNAL	1	20			
FT	CHAIN	21	501		EXTERIOR MEMBRANE GLYCOPROTEIN.	
FT	CHAIN	502	851		TRANSMEMBRANE GLYCOPROTEIN.	
FT	CARBOHYD	37			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	70			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	114			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	127			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	134			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	142			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	157			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	184			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	195			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	227			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	230			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	261			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	267			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	278			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	289			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	299			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	355			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	361			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	388			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	398			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	401			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	438			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	453			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	456			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	601			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	610			N-LINKED (GLCNAC . .) (POTENTIAL.)	
FT	CARBOHYD	626			N-LINKED (GLCNAC . .) (POTENTIAL.)	
QO	SEQUENCE	851 AA; 97178 MW; 3B02FCDD0B5FF118 CRC64;				

	Query March	2.5% Best Local Similarity	Score 100, Pred. No. 21, Matches 105, Conservative	DB 1, Length 851, Mismatches 185, Indels 160, Gaps 29,
Qy	292 GSLFLR---ELLIEFGTDHYGADTPTNEMPPS-----SEPSY-----LAAAT 331	167 GTCMRHGNISVILESCKHYWDMKFRYCAPPGFALLRCDDTNYSGFEPKSKYVAASC 246		
Db	332 TAYEAMTAVDTEAVWL-LOGTFLFOHPO-FWGAQIRAVLG-----AVPRRL 378	247 TRMETQT-----STWGFNGTRAENRTYIYHCKDKRITIIISLKYNYLTHCKRPGNKT 301		
Qy	379 LV-LDLA---ESQPYTRTASQGGQPFIMCMHNGCNGLFGALVAVNGSPEARLF 433	302 VPIITLISGRFHRFRPYNNK---PGQMCW-----FOGNV-----IEANFEVKOTLAKH 348		
Db	434 P-----NSTWYGTGMAPEGISQNEVVS-----LMAELQMRXD PVDLDAW 474	349 PRYGTNDTGKINTKFGISDP-----EVTYMTNCRGSFLVCNMTW-----PLNW 395		
Qy	475 VTSFAARRYGVSHP---DAGAARLLRSYV--NCSGE-ACRGHNSPLVRRPSLOMT 527	396 VENKTQTHGVACPIRQIINTMHKGTINYVLLPRREGELTGNSTVSIINIDSDGNOT 455		
Db	528 SIWNRSDDPEAMALLTSAPSLATSAFRDILLDRANQELVSLYEBA-----RSA 582	456 NITTS-REVAELVLYLEGD-----YKLLVETIPPAPTKERYSSAPVANKRGV 503		

QY 583 YLSKELASLLRAGVLAPELLPALBVLASDSRFLGSLWLEQARAAVSEADPEY-ON 641  
 DB 504 FVLGFGFLATNAGSANGASL-----TISAQRTLLAGIVQO-----QQCLLDVVKRQO 552  
 QY 642 SRVQLTLMP-----EGNILDYANKOLAGL-----VANYTTPMR--- 676  
 DB 553 EMLRLTWGTRKQLQAVNTAIEKYLKQAOQNSMGCAFQVCHTTPVWNVDSLTPEMNMNT 612  
 QY 677 -----LFLPALVDSVAGIPIFOQHOPDKVFOLEQ 706  
 DB 613 WOEWKRVHYLEA---NISQSLQAOIQOEKNMYELOK 647

RESULT 15  
 NME4\_MOUSE STANDARD; PRT; 1323 AA.  
 ID NME4\_MOUSE  
 AC 003391;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glutamate (NMDA) receptor subunit epsilon 4 precursor (N-methyl)  
 GN D-separate receptor subtype 2D (NR2D) (NMDAR2D).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid:10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA MEDLINE=93050214; PubMed=1365220;  
 RA Ikeda K., Nagaawa M., Mori H., Araki K., Sakimura K., Watanabe M.,  
 RA Inoue Y., Mishina M.;  
 RT "Cloning and expression of the epsilon 4 subunit of the NMDA receptor  
 RT channel";  
 RL FEBS Lett. 313:34-38(1992).  
 RN [2]  
 RP REVISIONS.  
 RA Ikeda K., Nagaawa M., Mori H., Araki K., Sakimura K., Watanabe M.,  
 RA Inoue Y., Mishina M.;  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ database.  
 CC -!- FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS  
 CC POSSESSES HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT  
 CC SENSITIVITY TO MAGNESIUM AND IS MEDIATED BY GLYCINE.  
 CC -!- SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D12822; BAA0254.1; -  
 DR HSSP: P19491; IGR2.  
 DR MGD: MGJ:95823; Grind.  
 DR InterPro: IPR001828; ANF\_receptor.  
 DR InterPro: IPR001330; Ion\_glu\_receptor.  
 DR InterPro: IPR001508; NMDA\_receptor.  
 DR InterPro: IPR001311; SBP\_glu\_receptor.  
 DR Pfam: PF01094; ANF\_receptor; 1.  
 DR Pfam: PF00060; 1lg\_chan; 1.  
 DR PRINTS: PR00177; NMDARECEPTOR.  
 DR SMART: SM00079; pape; 1.  
 KM Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;  
 KM Ionic channel; Magnesium.  
 FT SIGNAL 1 27  
 FT CHAIN 28 1323  
 FT DOMAIN 28 580  
 FT EXTRACELLULAR (POTENTIAL).  
 FT POTENTIAL. RECEPTOR SUBUNIT  
 FT EPSILON 4.  
 FT EXTRACELLULAR (POTENTIAL).

FT TRANSEM 581 601  
 FT DOMAIN 602 623  
 FT TRANSEM 624 644  
 FT DOMAIN 645 653  
 FT TRANSEM 654 674  
 FT DOMAIN 675 681  
 FT TRANSEM 842 862  
 FT DOMAIN 863 1323  
 FT DOMAIN 278 283  
 FT DOMAIN 905 913  
 FT DOMAIN 1030 1035  
 FT DOMAIN 1197 1201  
 FT SITE 639 639

FT CARBOHYD 89 89  
 FT CARBOHYD 349 349  
 FT CARBOHYD 363 363  
 FT CARBOHYD 464 464  
 FT CARBOHYD 566 566  
 SQ SEQUENCE 1323 AA; 142907 MM; 8AE9878F90DD0921 CRC64;

Query Match 2.58; Score 100; DB 1; Length 1323;  
 Best Local Similarity 22.78; Pred. No. 39;  
 Matches 106; Conservative 46; Mismatches 156; Indels 158; Gaps 28;

QY 162 INITIALMSGGEATWQRYVYALGLTQAEINEFFTGPAFLAMGKGNLH-----TWGGLP 215  
 DB 48 LNVALVFSGP-----AYAAEARL-----GPAVAAVRSPLGDIVRVALVLNNSDP 93  
 QY 216 PSWHIKQLYIQ-----HRYL--DQNRSGMTPVLPAPAGHYPEAVTRYFPQVNT 263  
 DB 94 -----RSLVQLCDLLSLGLRHVGVEPDSRAVAVPIIDFLSAGTSLPIVAHG----- 143  
 QY 264 KMGSMGHFNGSYSCSFLAPEDPIPIGLSLPRLIKERTGHTHYGADTFNEMO----- 318  
 DB 144 -----GALVLTPEK-----GSTFLQ-----LGSIEQLQVFEV 175  
 QY 319 -PSSSEPSYLAATT-----VYEMTAVDTEAVMLLOGWLFQHPQF-WGPAQIRAVLGA 372  
 DB 176 LEEYDWTSPVAVTTRAPGRHRAFLSYIEVLTDG--SLVGW--EHRGALLDPGAGEAVLGA 231  
 QY 373 VPRG-----RLVLVDLFAESQPIY--TRTASFOGQPIPIWCMH--NFGNHGLFG-ALE 421  
 DB 232 QLRVSQAQIRLLFC-AREEAEPIVRAEEAGLTPGVWPMVBPQLAGGGSGSGVPGPPL 290  
 QY 422 AVNGGPEARLFPNSTWGTGMAPEGISQNEVYSLMAELGWRKDPVPDLAAVTSFA-- 479  
 DB 291 LPGAFLPAGLF-----AVRSAGWRDDLARRVAAGVAVVARG 327  
 QY 480 ---ARRYGVSHPDAGAMRLLIRSVYNGSGEACRGHNSPLVRRPSLQ--NNTSIWYN 532  
 DB 328 AQAALLRDYGL-DEL-----GHGCRFQNRTH--RGESLHRYFVNIT-WDN 368  
 QY 533 RSDVFEAMRLLTSPSLATSPAFRYDLDTQAOVELSLYEE 578  
 DB 369 RDSYFNEDGFLVN--PSLV-----VISLTDRTWEVVGSMEOQ 404

Search completed: February 13, 2004, 16:15:50  
 Job time : 26 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2004, 15:59:34 ; Search time 61 Seconds

(Without alignments)  
3143.165 Million cell updates/sec

Title: US-09-836-613-2

Perfect score: 3939  
Sequence: 1 MEAVVAAGVLLAGAG.....VDLAKKIFLKYPGWAGSW 743

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mnc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriopl:\*  
17: sp\_archaeopl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3306	83.9	639	4 Q14769	Q14769 Homo sapien
2	3305	83.9	739	11 O54752	O54752 mus musculu
3	3302	83.8	739	11 O88325	O88325 mus musculu
4	2613.5	66.3	753	13 Q90276	Q90276 dromaius no
5	2611.5	66.3	753	13 Q90275	Q90275 dromaius no
6	1522.5	38.7	806	10 Q9FN43	Q9FN43 arabidopsis
7	1445	36.7	811	10 Q9ZRA5	Q9ZRA5 nicotiana t
8	1433.5	36.4	778	5 Q9VLL5	Q9VLL5 diosiphila
9	1060.5	26.9	770	16 Q9AAQ6	Q9AAQ6 caulobacter
10	1027.5	26.1	715	5 Q9NAP6	Q9NAP6 caenorhabdi
11	962	24.4	798	16 Q8P87	Q8P87 xanthomonas
12	854	21.7	2104	16 Q8XM24	Q8XM24 clostridium
13	143.5	3.6	1862	2 Q93TW7	Q93TW7 elismateila
14	128	3.2	444	3 Q96W93	Q96W93 gibberella
15	124.5	3.2	783	16 Q9FBV4	Q9FBV4 streptomyces
16	122.5	3.1	3247	12 Q65553	Q65553 bovine hept

17	121	3.1	2116	2 Q93N89	Q93N89 streptomyces
18	121	3.1	4342	16 Q91157	Q91157 pseudomonas
19	121	3.1	4472	2 O33954	O33954 streptomyces
20	120	3.0	3729	2 O33956	O33956 streptomyces
21	118	3.0	2785	17 Q8ZYB9	Q8ZYB9 pyrobaculum
22	117.5	3.0	1490	17 Q8PM14	Q8PM14 methanococcus
23	117	3.0	1374	17 Q9YCP5	Q9YCP5 aeropyrum p
24	116	2.9	1799	2 Q8KPM5	Q8KPM5 streptomyces
25	115.5	2.9	973	16 Q86728	Q86728 streptomyces
26	115.5	2.9	1319	4 Q9HCD3	Q9HCD3 homo sapien
27	115	2.9	725	16 Q9A722	Q9A722 caulobacter
28	115	2.9	7257	2 Q9K127	Q9K127 polyangium
29	114.5	2.9	2338	16 Q8XT19	Q8XT19 ralsionia s
30	114	2.9	479	2 Q86001	Q86001 springmonas
31	113.5	2.9	1030	16 Q8P685	Q8P685 xanthomonas
32	113.5	2.9	2164	2 Q93N86	Q93N86 streptomyces
33	113	2.9	1281	16 Q91211	Q91211 pseudomonas
34	113	2.9	7257	2 Q916C7	Q916C7 polyangium
35	112.5	2.9	380	16 Q8P1D5	Q8P1D5 xanthomonas
36	112.5	2.9	839	16 Q8P3W4	Q8P3W4 xanthomonas
37	112	2.8	596	2 Q9R8J7	Q9R8J7 xanthomonas
38	112	2.8	768	3 Q8J2T4	Q8J2T4 grifolia fro
39	111.5	2.8	484	10 Q9FFZ1	Q9FFZ1 arabidopsis
40	111.5	2.8	689	10 Q9AWX9	Q9AWX9 oryza sativ
41	111.5	2.8	2205	12 Q99FK2	Q99FK2 porcine tes
42	111	2.8	1027	16 Q9AAQ1	Q9AAQ1 caulobacter
43	111	2.8	10917	2 Q93NM6	Q93NM6 streptomyces
44	110.5	2.8	582	2 Q917A1	Q917A1 streptomyces
45	110.5	2.8	614	2 Q916K3	Q916K3 mycobacteri

## ALIGNMENTS

RESULT 1  
ID Q14769 PRELIMINARY; PRT; 639 AA.  
AC Q14769;  
DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (T-EMBLrel. 08, Last annotation update)  
DE Alpha-N-acetylglucosaminidase.  
GN UPHSD2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96359157; PubMed=8703123;  
RA Zhao Z., Yazdani A., Shen Y., Sun Z., Bailey J., Caskey C.T.,  
Lee C.C.;  
RT "Molecular dissection of a cosmid from a gene-rich region in 17q21 and  
RT characterization of a candidate gene for alpha-N-acetylglucosaminidase  
RT with two cDNA isoforms".  
RL Mamm. Genome 7:686-690(1996).  
DR EMBL; U78465; AAB36505.1; -.  
SQ SEQUENCE 639 AA; 72027 MW; 21847A1A2E978B6 CRC64;

Query Match 83.9%; Score 3306; DB 4; Length 639;  
Best Local Similarity 99.7%; Pred. No. 3.5e-233;  
Matches 617; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 125 TPNRYRYONVCTQSYSPVWMDWAEERIDWALNGINLALWAGQEAIRVYIAGL 184  
DB 21 TARRRYRYONVCTQSYSPVWMDWAEERIDWALNGINLALWAGQEAIRVYIAGL 80  
QY 185 TQAEINEFTGPAFLAMGNGNLTHTMDGFLPSSWHTIKOYLQHRVLDQNRSGMTPVPA 244  
DB 81 TQAEINEFTGPAFLAMGNGNLTHTMDGFLPSSWHTIKOYLQHRVLDQNRSGMTPVPA 140  
QY 245 FGHVPEAVTRVFPQVNTYKMGWGHFNCSYSCFLAPEDDIPFIIGSLFLRELKBERG 304

```

Db      141 FAGHPEAVTRVPEQVNVITMGSGWGHFNCSYSCSFLAPEDPIFPIIGSIFLNEIKKEFG 200
Qy      305 TDHIYGADTNEHQPPSSPSYLAATACTAYEAMTAVDTAVVLLQGMFLFOHOPQMGPA 364
Db      201 TDHIYGADTNEHQPPSSPSYLAATACTAYEAMTAVDTAVVLLQGMFLFOHOPQMGPA 260
Qy      365 QIRAVLCAVPRGRLLVLDLPAESQPVYTRTASFOGQPFIVCMILHFGNHGLFGALEAVN 424
Db      261 QIRAVLCAVPRGRLLVLDLPAESQPVYTRTASFOGQPFIVCMILHFGNHGLFGALEAVN 320
Qy      425 GGPEAAFLFNSITWVGMAPEGISONEVYYSIMAEIGMRKDPVPLAAVWTSFAARVY 484
Db      321 GGPEAAFLFNSITWVGMAPEGISONEVYYSIMAEIGMRKDPVPLAAVWTSFAARVY 380
Qy      485 VSHPDAGAAARLLLSRYNCSGEACRGNHNSPLVRPSLQMTSITWYNSDYVEAARLL 544
Db      381 VSHPDAGAAARLLLSRYNCSGEACRGNHNSPLVRPSLQMTSITWYNSDYVEAARLL 440
Qy      545 TSAPSLATSPAFRDLDLTDROAVQELVSLYEBAASAVLSKEIASLLRAAGVLAELLP 604
Db      441 TSAPSLATSPAFRDLDLTDROAVQELVSLYEBAASAVLSKEIASLLRAAGVLAELLP 500
Qy      605 ALDEVILASDSRFLIGSWLEQARAAVSEAEADFEQNSRYQLTLMGPEGNIIDYANKOLA 664
Db      501 ALDEVILASDSRFLIGSWLEQARAAVSEAEADFEQNSRYQLTLMGPEGNIIDYANKOLA 560
Qy      665 GLVANYTTPWRRLFLELVDSVAQGIFFQOHQDKVFLQLEQAFVLSKQRYPSQPRGDTV 724
Db      561 GLVANYTTPWRRLFLELVDSVAQGIFFQOHQDKVFLQLEQAFVLSKQRYPSQPRGDTV 620
Qy      725 DLAKKIFLKYYPGWVAGSW 743
Db      621 DLAKKIFLKYYPGWVAGSW 639

RESULT 2
054752 PRELIMINARY; PRT; 739 AA.
ID 054752;
AC 054752;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Naglu (EC 3.2.1.50).
GN Naglu.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBAF1J; TISSUE=Liver;
RA Zhao K.W., Li H.H., Neufeld E.F.;
RT "Cloning and expression of Mouse gene encoding the lysosomal alpha-N-
RT acetylglucosaminidase.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U85247; AAB88084.1; -.
DR MGD; MGI:135164; Naglu.
DR InterPro; IPR02086; Aldenhyde dehydr.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
KM Glycosidase; Hydrolase.
SQ SEQUENCE 739 AA; 82610 MW; B414C336A04EFOEF CRC64;

Query Match 83.9%; Score 3305; DB 11; Length 739;
Best Local Similarity 83.0%; Pred. No. 5,2e-233;
Matches 611; Conservative 53; Mismatches 70; Indels 2; Gaps 1;

```

```

Qy      121 LTEATPNRYRYQNVCTQSYSEFWWMDWAREREIDMNLGINIALAMSGEALIMQRYVL 180
Db      119 LTEATPNRYRYQNVCTQSYSEFWWMDWAREREIDMNLGINIALAMSGEALIMQRYVL 178
Qy      181 ALGLTQAEINFEFFGPAFLANGWGNHTWDGFLPESWHIKQLYLQHRVLDQNSFGMT 240
Db      179 ALGLTQAEINFEFFGPAFLANGWGNHTWDGFLPESWHIKQLYLQHRVLDQNSFGMT 238
Qy      241 VLPAPAGVPAVRVFPQVNVITMGSGWGHFNCSYSCSFLAPEDPIFPIIGSIFLNEIKKEFG 300
Db      239 VLPAPAGVPAVRVFPQVNVITMGSGWGHFNCSYSCSFLAPEDPIFPIIGSIFLNEIKKEFG 298
Qy      301 KEFGTDHIYGADTNEHQPPSSPSYLAATACTAYEAMTAVDTAVVLLQGMFLFOHOPQ 360
Db      299 KEFGTDHIYGADTNEHQPPSSPSYLAATACTAYEAMTAVDTAVVLLQGMFLFOHOPQ 358
Qy      361 MGPAQIRAVLCAVPRGRLLVLDLPAESQPVYTRTASFOGQPFIVCMILHFGNHGLFGAL 420
Db      359 MGPAQIRAVLCAVPRGRLLVLDLPAESQPVYTRTASFOGQPFIVCMILHFGNHGLFGAL 418
Qy      421 EAVNGGPEAAFLFNSITWVGMAPEGISONEVYYSIMAEIGMRKDPVPLAAVWTSFAA 480
Db      419 EAVNGGPEAAFLFNSITWVGMAPEGISONEVYYSIMAEIGMRKDPVPLAAVWTSFAA 478
Qy      481 RRYGVSHPDAGAAARLLLSRYNCSGEACRGNHNSPLVRPSLQMTSITWYNSDYVEAAR 540
Db      479 RRYGVSHPDAGAAARLLLSRYNCSGEACRGNHNSPLVRPSLQMTSITWYNSDYVEAAR 538
Qy      541 RLTLTAPSLATSPAFRDLDLTDROAVQELVSLYEBAASAVLSKEIASLLRAAGVLAELLP 600
Db      539 RLTLTAPSLATSPAFRDLDLTDROAVQELVSLYEBAASAVLSKEIASLLRAAGVLAELLP 598
Qy      601 ELLPDLDEVILASDSRFLIGSWLEQARAAVSEAEADFEQNSRYQLTLMGPEGNIIDYANK 660
Db      599 ELLPDLDEVILASDSRFLIGSWLEQARAAVSEAEADFEQNSRYQLTLMGPEGNIIDYANK 658
Qy      661 KQLAGLVANYTTPWRRLFLELVDSVAQGIFFQOHQDKVFLQLEQAFVLSKQRYPSQPR 720
Db      659 KQLAGLVANYTTPWRRLFLELVDSVAQGIFFQOHQDKVFLQLEQAFVLSKQRYPSQPR 718
Qy      721 GDTVDLAKKIFLKYY 736
Db      719 GDTVDLAKKIFLKYY 734

RESULT 3
088325 PRELIMINARY; PRT; 739 AA.
ID 088325;
AC 088325;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Alpha-N-acetylglucosaminidase.
GN Naglu.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129SV;
RC MEDLINE=20056274; PubMed=10588735;
RA Li H.H., Yu W.H., Rozenfurt N., Zhao H.Z., Lyons K.M.,
RA Anagnostaras S., Farnsworth M.S., Suzuki K., Vanter M.T., Neufeld E.F.;
RT "Mouse model of Sanfilippo syndrome type B produced by targeted
RT disruption of the gene encoding alpha-N-acetylglucosaminidase.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:14505-14510(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVevtaCBr; TISSUE=Spleen;
RC Gloeckner C.J., Brechtling R., Moeller G., Adamski J.;
RA "Characterization of HSD17B1 gene in mice.";

```

Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF003255; AAC26842.1; -  
 DR EMBL: AF363242; AAM2194.1; -  
 DR MGI: 1351641; Naglu.  
 DR InterPro: IPR002086; Aldenhyde dehydr.  
 DR PROSITE: PS00070; ALDENHYDE DEHYDR. CY5; 1.  
 DR SEQUENCE 739 AA; 82596 MW; 01C71966557CDE27 CRC64;

Query Match 83.8%; Score 3302; DB 11; Length 739;  
 Best Local Similarity 82.9%; Pred. No. 8.6e-233;  
 Matches 610; Conservative 54; Mismatches 70; Indels 2; Gaps 1;

QY 1 MEAVAAVAVGVLILAGAGAADEARAAVAVLVARLGGPAAADPSVSVERALAKP 60  
 DB 1 MEAGLAVILGFLIL--AGGSVGEARAKAKAVRELVRLLGPGRAAFVSVERALDES 58  
 QY 61 GLDITSLGGGGAARVRGSGTGVAAAAGLHRYLDFCGCHVANGSGQLRLPRPLAVGE 120  
 DB 59 GLDITSLGGGGAARVRGSGTGVAAAAGLHRYLDFCGCHVANGSGQLRLPRPLAVPDG 118  
 QY 121 LFEATPNRYRYONCTQSYSPFMMWMAWREIEMALNGINLALMSGGEALWGRVYL 180  
 DB 119 LFEATPNRYRYONCTQSYSPFMMWMAWREIEMALNGINLALMSGGEALWGRVYL 178  
 QY 181 ALGELTOAEINEFTGPAFLANGMGNLHTWDGFLPPSWHLKQLYLQHRVLDQMSFGMT 240  
 DB 179 ALGELTQSEIDITYFGPAFLANGMGNLHTWDGFLPPSWHLQYVLRLLDRMSFGMIP 238  
 QY 241 VLPAFAGVPAARVTRVFQVNVTKMGSNGHNCSCSGLLAPEDPLPFIIGSLFLBELT 300  
 DB 239 VLPAFAGVPAARVTRVFQVNVTKMGSNGHNCSCSGLLAPEDPLPFIIGSLFLBELT 298  
 QY 301 KEFGDTHYAGDTENEMOPSSSEPSYLAATTAYEAMTAVDTAVALWLLQGMFLFOHOP 360  
 DB 299 KEFGDTHYAGDTENEMOPSSSEPSYLAATTAYEAMTAVDTAVALWLLQGMFLFOHOP 358  
 QY 361 WGPQIRAVLCAVPRGRLLVLDLFAESQPVYTRTASFOGQPFVWCMHNFQNGHGLFGL 420  
 DB 359 WGPQIRAVLCAVPRGRLLVLDLFAESQPVYTRTASFOGQPFVWCMHNFQNGHGLFGL 418  
 QY 421 EAVNGGPAARLFPNSTWGTGMAPEGISONEVYVSLMAELGMRKDPVLDLAAMVTSFAA 480  
 DB 419 EDVNRGPAARLFPNSTWGTGMAPEGISONEVYVSLMAELGMRKDPVLDLAAMVTSFAA 478  
 QY 481 RRYGVSHDAGAAARLLRSVYNGSGEACRGNHNSPLVRBPSLQNTSITWNSDVEAM 540  
 DB 479 RRYGVSHDAGAAARLLRSVYNGSGEACRGNHNSPLVRBPSLQNTSITWNSDVEAM 538  
 QY 541 RLILTSASLATSAPFRYDLDTROAVQELVSLYEERASATSKELASILPAGVILAY 600  
 DB 539 RLILTSASLATSAPFRYDLDTROAVQELVSLYEERASATSKELASILPAGVILAY 598  
 QY 601 ELPLALDEVLASDSRFLGSLWLEOARAAVSEAEADFEONSRYQLTLWGEGNILLDAN 660  
 DB 599 ELPLALDEVLASDSRFLGSLWLEOARAAVSEAEADFEONSRYQLTLWGEGNILLDAN 658  
 QY 661 KQLAGLVANVTTPKRRLEFLVDSVAGGIFPQOHQPDKNVFOLEAFVLSKQRYPSQPR 720  
 DB 659 KQLAGLVANVTTPKRRLEFLVDSVAGGIFPQOHQPDKNVFOLEAFVLSKQRYPSQPR 718  
 QY 721 GDTVDLAKKIFLKYP 736  
 DB 719 GDTVDLAKKIFLKYP 734

RESULT 4

Q90276 PRELIMINARY; PRT; 753 AA.

AC Q90276; 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)

DE Lysosomal alpha-N-acetyl glucosaminidase.

OS Dromaius novaehollandiae (Emu).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Ornithura; Aves; Palaeognathae; Casuariiformes; Dromiidae;

OC Dromaius;

OX NCBI\_TaxID=8790;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=2109063; PubMed=11414757;

RA Aronovich E.L.; Johnson J.M.; Wang P.; Giger U.; Whitley C.B.;

RT "Molecular Basis of Mucopolysaccharidosis Type IIb in Emu (Dromaius

novae-hollandiae): An Avian Model of Sanfilippo Syndrome Type B.";

RL Genomics 74:299-305(2001).

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

DR EMBL: AF31668; AAK73654.1; -

Q90Z75		
ID	Q90Z75	PRELIMINARY;
AC	Q90Z75;	PRT; 753 AA.
DT	01-DEC-2001 (TREMBlrel, 19, Created)	
DT	01-DEC-2001 (TREMBlrel, 19, Last sequence update)	
DT	01-OCT-2002 (TREMBlrel, 22, Last annotation update)	
DE	Lysosomal alpha-N-acetyl glucosaminidase.	
OS	Dromaius novaehollandiae (Emu).	
OC	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Archosauria; Aves; Palaeognathae; Casuariiformes; Dromadidae;	
OC	Dromaius	
CX	NCBI_TaxID=8790;	
RN	[1]	
RN	SEQUENCE FROM N.A.	
RX	MEDLINE=21309063; PubMed=11414757;	
RA	Acronovich E.L., Johnston J.M., Wang P., Giger U., Whitley C.B.;	
RT	"Molecular Basis of Mucopolysaccharidosis Type IIIB in Emu (Dromaius	
RT	novaehollandiae): An Avian Model of Sanfilippo Syndrome Type B.",	
RL	Genomics 74:299-305(2001).	
DR	EMBL; AF31669; AAK73653.1; --	
DR	SEQUENCE 753 AA; 84008 MW; 3BD04AEE30B49A85 CRC64;	

	Query Marc3	66.3%	Score 2611.5	DB 13	Length 753
	Best Local Similarity	64.7%	Pred. 2.8e-182		
	Matches 489	Conservative 104	Mismatches 145	Indels 13	Gaps 6
QY	4 VAVAAAVGVLLLAGAGC-----AAGDEAFBAAVRALVARLIGQGPAADEVSYERAL	56			
Db	5 VRLTALAAWMAAARAPSPPLPYVRAGADAAGZAAVALRLRLLGPPAAA-VALSDVGL	63			
QY	57 AAKPGLDITYSLGG--CGAARYVRGSGTGVAAAGGHRRLRDPGCGAHYMSGSLRPRPL	114			
Db	64 AA-GGLDITRYASPPRAAVALVAASSGVAAGGHRRLRDLCCGHSWGGRRLRPDDL	122			
QY	115 PAVPGELTEATPNRYRYVYQVCTGYSFVMDWAPREREIDWALINGINIALAWSQEA	174			
Db	123 PRYPARIRATAPRFRFYQVCTGYSFAMDMWAMEEIDWALSGINIALAFAQGEAV	182			
QY	175 MCVVYALALGTOAEINEEFTGPAPLAMPGRMGLHTWGPDPSPSHIKQLTQHRVLDOR	234			
Db	183 MCVYVYSLGINSQSEIDEIFYTGAPFLANRNGNLHGMAGPLPRANHLQLYQVRYLERK	242			
QY	235 SFQMTPLVLPAGHVEAVTRVPQVNYTKMGSGWGHFNCYSQSFLLAPEDPIFPIISL	294			
Db	243 SLGMITVLPFAGHVDPQVLRAPPRVYNAIRLGGMSHEPCTSYLLDPEDPEFQVIGTL	302			
QY	295 FLKELLKERCTDITVADTFENMQPSSPSYLAATTAYAEANTANDTAVALWLGWLF	354			
Db	303 FLKELLKERCTDITVADTFENMNPSSPALTSVSAVRSHTGADPRAYVALMGWLF	362			
QY	355 QHQPQFQWPAQIAVAVGAVPRGRLLVLDLFAESQPVYTRTASFQGPFIICMLHNGGNH	414			
Db	363 QHQPDPQAPQAVRALHGVPLGRMIVLDFABSRPVYQWTSFYGQPFICMLHNGGNH	422			
QY	415 GLGALFAVNGGEARLFPNSTMTGTCGAPRGISQNEVYSLAAELGMKKDVPYDLAAW	474			
Db	423 GLGFTVEALNHGFAPRRPNSTMTGTVLPGELEQNMYVELNNEIGMQOEPL-DLPEA	481			
QY	475 VTSFAARARYVSHPDGAAMRLILRSYVNSGSEACRGNRSPVLRPSLQMTNISYWN	534			
Db	482 VARAABERYAPAAAAAAMXLLISVYNCIS-VCVNNNRSPVLRPSLMDTEVYWN	540			
QY	535 DYFAAMLLITSAPLSATSPARVYLLDLTRQAVELVSLYEEARARAYLSKELASILA	594			
Db	541 DYFAAMLLITSAQAEIGSSPTFGYDLADYTRQAQVLSFYLLIRAFQFSRSLPELTA	600			
QY	595 GGVLYAELLALDEVLASDSRFLGSMLEQAAVAASDAADPFQORSRYQLTLMGEGN	654			
Db	601 GGVLYVLDLDELDDGLLSHRLFLIGKMSLSARAAYATSBRAEQEILNABRQVTLMGNGN	660			
QY	655 ILDYANKQLAGLVANYYTPRMRLFELAVDSYAQGIPIQGHQFDKAYFQLEQAFLYSQR	714			
Db	661 ILDYANKQLAGLVAVGYWMSLFYSALVESINSGSPHQDFQAVFQVERGIVNKR	720			

```

QY      715 YPSQPRGDTVDLAKKIFLKYYP 736
          ||:::|||||
Db      721 YPTAPVGDTLEISKIFLKYYP 742

```

RESULT 6	Q9FNA3	PRELIMINARY;	PR;	806 AA.
1D	Q9FNA3			
AC	Q9FNA3			
DT	01-MAR-2001 (TrEMBLrel. 16, Created)			
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)			
DE	Putative alpha-N-acetylglucosaminidase.			
GN	ATG513690.			
OC	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryotes; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eustroids II; Brassicales; Brassicaceae; Arabidops.			
OX	NCBI_Taxid=3702;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=Columbia;			
RC	MEDLINE=98069011; PubMed=9405937.			
EX	Kotani H., Nakamura Y., Sato S., Kaneko T., Aamizu E., Miyajima N.,			
RA	Tabata S.;			
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. II.			
RT	Sequence features of the regions of 1,044,062 bp covered by thirteen			
RT	physically assigned pl clones.";			
RL	DNA Res. 4:291-300(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Yanada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,			
RA	Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,			
RA	Carinini P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,			
RA	Karlin A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,			
RA	Mitani M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,			
RA	Seki M., Shimn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,			
RA	Theologis A.;			
RT	"Arabidopsis Full Length cDNA Clones";			
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Yanada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,			
RA	Deng J.M., Goldsmith A.D., Lee J.M., Oondara C.S., Quach H.L.,			
RA	Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,			
RA	Carinini P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,			
RA	Karlin A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,			
RA	Mitani M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,			
RA	Seki M., Shimn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,			
RA	Theologis A.;			
RT	"Arabidopsis Open Reading Frame (ORF) Clones";			
RL	Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AB006704; BAB08656.1; -			
DR	EMBL; AY080811; AAB87251.1; -			
DR	EMBL; AY117179; AAMS1254.1; -			
SC	SEQUENCE 806 AA; 92689 MW; 8F8500032F3BCFA2 CRC64;			

```

Query Match 38.7%; Score 1522.5; DB 10; Length 806;
Best Local Similarity 36.8%; Pred. 1.4e-102;
Matches 304; Conservative 161; Mismatches 223; Indels 95; Gaps 17.

QY 28 EAAAVRALVAELTGGPAADESVSVERALAAKPG-----LDTSLGGGAARVRVG 79
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 45 QESAAGKLGLOLL-----PTHQSPELRITISKDAGGTSCEVIERYDGPRIPEILIK 99
QY 80 STGVAAAGLHRYLEDFGCGYAV---SGSGL-RLEPR--LPAYGE-LTEATPRRYRY 132
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 100 TTGVEIASGLHMYLTKCNAAHVSMDKTGGIIVASVPGCHLPRIDSKRIETRRPFPNNY 159
QY 133 QNVCTQSSFFVMMWMAWRERIDWVNLNGVNLALAMSGCAIWMRVYALGLOLEINEF 192
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 160 QNVVTSSSYVWVGWGERERETIDWALGGINPLAFTQCEALIQCVKFRFPINSKEDLDY 219

```





AC Q9A06: 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Alpha-N-acetylglucosaminidase.  
 GN CC0540.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;  
 OC Caulobacteraceae; Caulobacter.  
 CX NCBI\_TaxId=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Mierman W.C., Feldblum T.V., Laud M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen U., Heidelberg J.F., Alley M.R.K., Onta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kojanay J.F., Smit J., Craven M.B., Khouri H., Shetty K.,  
 RA Uetreck T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,  
 RT "Complete genome sequence of *Caulobacter crescentus*."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 DR EMBL; AE005727; AAK2527.1; -  
 DR TIGR; CC0540; -  
 KW Complete proteome.  
 SQ SEQUENCE 770 AA; 84287 MW; 94A694F41C18D20 CRC64;

Query Match 26.9%; Score 1060.5; DB 16; Length 770;  
 Best Local Similarity 33.0%; Pred. No. 8.1e-69;  
 Matches 260; Conservative 132; Mismatches 302; Indels 93; Gaps 19;

QY 2 EAVVAAGVLLAGAGAGDEAREAAVRAIARLILBGPAPDSVSEALAKPG 61  
 DB 11 QAIISLAAAFAPASALAAAGTDG---VAAARASLKLPERRLAG-----AHLTYTFG 59  
 QY 62 LDT--YSLGGGAARVVRGSGTGVAAAAGLHRYLRDPCGCHVAVSGSQLRPPLPAVPG 119  
 DB 60 AERGWYAIIGKGA-ISISGSPVALVRGAVALRQGLAHVSEGDVAVQAGVPAAG 118  
 QY 120 ELTATPRTYRYQNVCTQSISFVWMDMARWERIDIMMALNGINTALAMSGEALIMORY 119  
 DB 119 ARVE-TDFRHRAYNTCTGYGTTTPWGMGRWREIDIMMAHAGIDMPLAMGEQEVWALW 177  
 QY 180 LALILTOAINEFTGPFLANGRMGNLHTWDGFLPPSMHKKOYLQHRVLDOVRSQMT 239  
 DB 178 REFGLSERELADYSGAFTPMHMGNIKGLKAPLPIAMIDKDLQKYLIGNRSLGNT 237  
 QY 240 PVLPAGHVEAVTRVFPQVNTKMGWGHFNCSYSCSFLLAEDDIPPIIGSLFREL 299  
 DB 238 PILPAFGYVPAFAEKPKKARIYRMPWEGFHETV---WLDPADLPFAKIAARFIATY 293  
 QY 300 IKKGTGTHIYAGDFFENMQP-----SSRPSV-----LA 328  
 DB 294 TETAGAGYVYADSFNEPLPIINDGADARDAAGDGTANTAVTKYEVDPALKAQRLA 353  
 QY 329 AATAVVEAMTAVTTEAVMLQGMFLFOHQFQWGAQIRAVLGAVERGRLLVLDFAESQ 388  
 DB 354 AYGAIVDSIRKTPDAVWVWQGLFGADSHFWPRAISAVLSLPVDDKMLIIDIGNDY 413  
 QY 389 P-VYTRTASFOGQPIFIMCLHNGNHLFGAL-----EAVNGGEAARLPFNSTWVG 440  
 DB 414 ENWKNAKAFGSKWIVGYVHNGSNVYVGDLPFYODIPALIAANDAKL-----AG 467  
 QY 441 TGMAPGISQNEVVYSLAELGMRKDPVPLDAAVTFFAARVYVSPDGAARLLRS 500  
 DB 468 FGMPEGLHNSIYEAAYDLAMSEGAAP-ATWLTYYARVYKTSFPALDALGQVLEA 526  
 QY 501 VYNGSGEACRGHNSP---LVRRPSIQMNTSIY--NRSDVFRAWLLITSAPSLATSP 554  
 DB 527 AFSTRYSPRMWKSAGAYLFFKRPATVGDFFQHPGRACLEAAVAVALTALAFYQGP 586  
 QY 555 AFRVLLDLTRQAVQELSLVYEEARSAVYLSKELASLLRAGGVLAAYELLPALDEVLASDS 614

DB 587 LFVLDLTATRHILATMKIDDLQVAVAAVRRGDTA---AGDARVEL-----EALASI 637  
 QY 615 RFLIG-----SWEQARAAAVSEAEADFEONSRYOITLMGPGNITLDYANKQIAGIV 667  
 DB 638 DKLIGVQPTLATWIDERAVAGDTPADAAAYAAKKAQVTITMGSGNINDYASKWQGLY 697  
 QY 668 ANYTPRRMLFLEALVDSVAGQIPQCHQFDKNVFOLEQAFVLSKORY---PSQPRDPT 723  
 DB 698 KSPYLPFRMSRFLDAL--KAAGTGTFDEVTYTRGVAMERAWAEAVARREKPADPGEI 755  
 QY 724 VDLAKKI 730  
 DB 756 KTLIARI 762

RESULT 10  
 ID Q9NAP6 PRELIMINARY; PRT: 715 AA.  
 AC Q9NAP6:  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Hypothetical protein K09E4.4.  
 GN K09E4.4.  
 OS *Caenorhabditis elegans*.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Felodetinae; Caenorhabditis.  
 CX NCBI\_TaxId=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smye R.;  
 RT Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode *C. elegans*: A platform for  
 RT investigating biology."  
 RL Science 282:2012-2018 (1998).  
 DR EMBL; Z83234; CAB0170.2; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 715 AA; 82640 MW; 73CE7FFFD05CD995 CRC64;

Query Match 26.1%; Score 1027.5; DB 5; Length 715;  
 Best Local Similarity 33.2%; Pred. No. 1.9e-66;  
 Matches 243; Conservative 123; Mismatches 312; Indels 53; Gaps 15;

QY 9 AVGVLLAGAGAGDEAREAAVRAIARLILBGPAPDSVSEALAKPG 59  
 DB 8 ALIIVLGA-----ERAGALRVQLDQDYDKNSMRSVLRVAFEWLEEFADVKLISAE 62  
 QY 60 PGLDITYSGGGGAARVRGSG-----TGVAAGLHRYLRDPCGCHVAVSGSQLRLP 111  
 DB 63 NG---FOLPKSLAQIETKNSGRKTVTANTPPDALNAINITYIKTCLQVSNSS--FS 117  
 QY 112 RPLPAVGEILTATENRYRYQNVCTQSISFVWMDMARWERIDIMMALNGINTALAMSGQ 171  
 DB 118 SCGRKRTSDFINFSKQIRYFNMCTFSYFAWMEWPOWERIDIMMALNGINTALAMSGQ 177  
 QY 172 EALIMQRYIALGLTOAINEFTGPFLANGRMGNLHTWDGFLPPSMHKKOYLQHRVL 231  
 DB 178 EILIMDIFGLGVQDELDYSFTSQAYLAMHMGMLKAYGGSLSDQVLMNDHNLKRIID 237  
 QY 232 QKRSFGMTFVLPAFGHVEAVTRVFPQVNTKMGWGHFNCSYSCSFLLAEDDIPPII 291  
 DB 238 RLLEIGTPIILTFAGFPVDHLETLFPAKSFRLLRMNFTSETSCMLSVSFDLFLFKI 297  
 QY 292 GGLFRELILKEFGTD--HIYGADTINEMQPPSS---EPSYLAATTAAYEAMTAVTNAV 346  
 DB 298 GSTFLRHQKMGFGGVNTWYSADPNEILPESAKFAKAYQOTQAQALNNSCKYDKNCV 357  
 QY 347 WLQGMFLFQHQGFQWGAQIRAVLGAVERGRLLVLDFAESQPVYTRTASFOGQPIFCM 406

```

Db      358 WVLQSGSFYDQ--WPMAIKSFLSAIPGNLLILDVAEVPAPQMTSFGGHHFVWCL 415
Qy      407 LHNFGNGHGLFGALBEANGGPEAARLFENSTWGTGMAPEGISQNEVYYSMAELGMRKD 466
Db      416 LHNFGSGRLRLKGLQIDKGYOLALMKAGSNLVGAGLSMEALDQYVYVYGFNIDRMWPE 475
Qy      467 PVPDLAAVTSFAARRYGVSHPDAGAAFLILRSYNSGEGACRGHNSPLYRPSLOMN 526
Db      476 PLP-LNNMLKAYSESRSYADPFVQAQKFTMLAGTFEKNQPEKWTGRFVFLYHRGFGRK 534
Qy      527 TSIWYRSVFEAMRELLTS-APSLATSPAFRYDLDLITROAVQELVSLYEEARSAVLS 585
Db      535 IETWFEVEETFSRFRLLPALVHTLGEHFLFRDLNDVREMTQ-----FEMGNELLS 588
Qy      586 KELASLL---RAGVLAAYELLPALDEVLASDSRFLGSMLEQARAAVSEADPYEQN 641
Db      589 MSBAFLMEDKQGA--SCEMLEMFQKLESYNSRDVQWIMENAKSIAPTSEERQVPEVT 646
Qy      642 SRQQLNPGEGNILDYANKQLAGVANYTTPRMFLFLALVDSVAQGIPIQGHQFQKV 701
Db      647 AGDILTWGPTQNDLYAHREWAGLMSGYGRWQF---CDWILEHDDENHTFEFSTSV 702
Qy      702 PQ-LEQAFVLS 711
Db      703 FPDVERPFSIS 713

RESULT 11
Q8PPH7  Q8PPH7  PRELIMINARY;  PRT;  798 AA.
AC      Q8PPH7;
DT      01-OCT-2002 (Tremblrel. 22, Created)
DT      01-OCT-2002 (Tremblrel. 22, Last sequence update)
DE      N-acetylglucosaminidase.
DE      XAC0709.
OS      Xanthomonas axonopodis (pv. citri).
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC      Xanthomonadaceae; Xanthomonas.
OX      NCBI_TaxID=92829;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=306 / ATCC 13902 / XV 101;
RX      MEDLINE=22022145; PubMed=12024217;
RA      da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
RA      Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA      Alves L.M.C., do Amaral A.K., Bertolini M.C., Camargo L.E.A.,
RA      Camarotte G., Canavan F., Cardozo J., Chambergo F., Clapina L.P.,
RA      Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA      Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA      Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA      Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA      Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA      Martins E.C., Meidanis U., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA      Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA      Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA      Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA      Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA      Setubal J.C., Kitajima J.P.;
RT      "Comparison of the genomes of two Xanthomonas pathogens with differing
RT      host specificities";
RL      Nature 417:459-463 (2002).
DR      EMBL; AB011700; AAM35598.1; -.
KW      Complete proteome.
SQ      SEQUENCE 798 AA; EC60943F5A84A7BE CRC64;

```

Query Match 24.4%; Score 962; DB 16; Length 798;  
 Best Local Similarity 32.1%; Pred. No. 1.4e-61;  
 Matches 251; Conservative 123; Mismatches 295; Indels 112; Gaps 23;  
 7 AAAGVLLLAGAGGAGDEAREAAVRLVRLGPPADPSVSVERALAAKPGIDTYS 66

```

Db      16 AVTLGALLATPVVLA-----ASPAQAVYQRLIG-ARAQPFMTV---AFRGDADWTR 65
Qy      67 LGGGGAARVRVRSSTGVAAGLHRYLRDFCGCHVMSGSQRLPRLPAV-FGELTEAT 125
Db      66 IDAGDT-VRIAGSGVALLARGAAYVILGGAASMEGGRVALPAQWPAVRSQV--RT 122
Qy      126 PNRYRYQNTCTQSYSTFWMDARMERETDMMNLNINLALMSGGEALIMQRYLALGLT 185
Db      123 PPARAVLNTCTGYTTPFDMPRMQREIDMHLHGDMLAEQGEALIMQALMREPDYV 182
Qy      186 CAEINEEFTGPAFLANGRMNLTWGDPLPSPWHIRQLYLQHRVLDQMSFGMTPLYLPF 245
Db      183 DDAIAEYFSGRATFPQMGKINIGYRAPLPQHIDSKVYLQKILRMELGNGQLPAPF 242
Qy      246 AGHVEAVTRVPPQVNTKMGSGHNCYSQSFLLAPEDPIPIIGSLFLRLIKEFGT 305
Db      243 AGYVPRAPQAHPAHAIYRRRAWECHETV---WLDPRDPLPAKLARRLLELYAQTYG 298
Qy      306 DHIYGADTFENMQP--SSEPSYLAAT-----TAVY 335
Db      299 GERYLADATFENMLPVAADDSDVAAPARYGDSINSDAAAKAVPPAQRDARLAEYQALY 358
Qy      336 EAMTAVDTEAVWLLQGLFQHPQFQWPAQIRAVIGAVPRGLVLIDLPAESOP-VYTRT 394
Db      359 RSLAQVVPKATVWQGLFQADRQFWQAQIAAFLGKVPDARLMLVIDIGNDRYFTGWKAS 418
Qy      395 ASFGQGPFIWCMLEHNGNGLFG-----ALEAVNGGEARLFPNSTWGTGMAPEG 447
Db      419 RAFPNGKMITGYVHNTGASNPFLYGDPAFYRHDIQALLADBDKXNL-----RFGVFPBG 472
Qy      448 ISQNEVYYSMAELGMRKQFVPDLAAVTSFAARRYGVSHPDAGAAFLILRSVNSGSE 507
Db      473 LHSNSVYELVYLAAM-EGHQSQSMQWLTHTYLARFGRSDAALLSAMSDLEAGIYQRYW 531
Qy      508 ACRGHRSP----LVRRPSIQMNTSIWYRSDFE-----AWRLLTAPSL 550
Db      532 SPRWMNRGAGAYLLFKRPT-----ADIVDDPRDQDRLRALDRLQANRY 580
Qy      551 ATPSPAFRYDLD-----LTQAVQELVSLYEEARSAVLSKELASLIRAGVLA--YELL 603
Db      581 ADAPLYYDILIEDARHLSLSQADROLAVY-----QAYDAGDPA---RGAQLARFTQLY 632
Qy      604 PALDEVLASDSRFLGSMLEQARAAVSEAE-ADFTQNSRYQLTLMGPGNILDYANKQ 662
Db      633 RGLD-ALVGGQHETLDWTCQAAAAGHDGLRRAYVGNARAVSWGGGDNLADYASKA 691
Qy      663 LAGLVANYTTPRMFLFLALVDSVAQGIPIQGHQFQKVQFLOEAF---VLSKQRYPSQ 718
Db      692 WQGYADIFYQRTFRFLSAIRAKKAGTPPDVAIVDHQLATWERSQADENSKQOPRE 751
Qy      719 P 719
Db      752 P 752

```

RESULT 12  
 Q8XM24 Q8XM24 PRELIMINARY; PRT; 2104 AA.  
 AC Q8XM24;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DE Probable alpha-N-acetylglucosaminidase.  
 DE CPE0866.  
 OS Clostridium peifringens.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1502;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=13 / Type A;  
 RX PubMed=11792842;  
 SH Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,

RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
 RT flesh-eater";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
 DR EMBL; AP003188; BAB80572.1; -  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR000421; PA58 C.  
 DR InterPro; IPR003861; FN\_III.  
 DR Pfam; PF00754; P5\_P8\_type\_C; 1.  
 DR Pfam; PF00041; fn3; 1.  
 DR SMART; SM00606; FN3; 1.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 2104 AA; 236010 MM; 5FE345659595ED7 CRC64;

Query Match 21.7%; Score 854; DB 16; Length 2104;  
 Best Local Similarity 28.6%; Pred. No. 4,5e-53;  
 Matches 214; Conservative 138; Mismatches 309; Indels 86; Gaps 23;

32 VRLVAVRLGGPAPDPSVERALAAKPGIDTSLGGGGAARVRVSGTGVAAAGLHR 91  
 208 IKNLVGRVIGEEFKDKTFEIRDO--NGNDVEVSDSDKVLKNGNVSLASGFNY 264  
 92 YLRDFCGCHVAMS--GSQRLRPRLPAVPELTPENRYRYQVNTOSYSFVWMDMA 148  
 265 YLKQY--CNVSYNPIMSGNLKMPETMPVGERVVIDPPEYRYALNCTSYTWSFWMD 322  
 149 RWEEDIMALNGINLALAMSGQALIMORVYLAGLQALNEPFTPAFLARMNLH 208  
 323 QYEEFLDMCANGVNLVDITIGSEVNLRTNERYGSDVEEKESGPAFAFWYQNMNT 382  
 209 TWGGLPSPWHIKQYLQHRVLDMQSRFGMTPLYPAFAHPEAVTVEFQVNTKMGV 268  
 383 GFGGLPMDWEQRLAKRKHDMQSGINPVLYGSGVPRPKENPBAQITSGGM 442  
 269 GHFNCSYSCFLPAA--EDPIPIIGSLFRLIKEFG-TDHYGADTFNEMQPPSS 322  
 443 ---CGFERPMLKTYVNEGEVDYFQNAVDVEKQKEVFCDVNFYGVDPFHE----- 492  
 323 EPSYLAANTTA-----VYE---AMTAVDTEAVWLLQGMFLQGPQFQWPAQIRAVLGA 372  
 493 ---GNTGDDNGKITYEITQNMKEHNDAYWVLIQNM-----QGNPSNNKEGL 538  
 373 VPRGRLVLDLPASQPYRTTASFOGPFIMCLHFGNGHGLFGALFAVNGGPEARL 432  
 539 TKDQAMVLDLFESEVSPDMRLEB-RDLPWIMNMLHFGNGMDAABEKL--ATEIPKA 595  
 433 FPNST-WGTGMAPEGISQNEVYYSIMAELEGKRDPPDLAAVTSFAARYGVSHDAG 491  
 596 LANEHWVGIGITPEALNTNPLAHELLFDMWSTBDQI-NFRTWEDYIERRYGKTKEIL 654  
 492 AAMRLLSVYVNGSGEACRGNRSPLYARPSLQW-NTSIV-----YNSDVEFAMRL 544  
 655 DAMNILLDTAKKNDYYQAAEISINARPGFGKASATWGHSHKTYVDKSGFEKALIFA 714  
 545 TSAISLATSPAFRDLDLTROAVOELVSYIEEARSAYLSKELASILRAGCVIAYLLP 604  
 715 KNYDEFKQSDAFLYDFADILKQLLANSAGEYEWCAVYNNNGEKEKRFVSGKE-LELIK 773  
 605 ALDEVTLASDSRFLGSMLEQARAASVSEAE--ADFYEONSRYOQLTWGPE-----GNILD 657  
 714 LQERVLSTRPEFLIGNWIEDARTMLKSDSDWTQKLPFENAAVLTWGSRRNADGGGLKD 833  
 658 YANQOLAGLVANYYTPRWLF--LEALVDSVAQGIIPQCHQFDKNV--FQLEQAFYLSK 712  
 834 YSNRQMSGLTEDDYYARWEKWINGLQAEILDGAKA-----FNIDWFMEYDWNKK 884  
 713 ---ORVSPQSGDTVDLAKKIFLKY 735  
 885 SDTDLKLYTEASNEVGLAKIAMEST 911

RESULT 13

Q93TW7  
 ID Q93TW7 PRELIMINARY; PRT; 1862 AA.  
 AC Q93TW7;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE MxAE.  
 GN MxAE.  
 OS Stigmatella aurantiaca.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;  
 OC Cyctobacterineae; Cyctobacteraceae; Stigmatella.  
 OX NCBI\_TaxId=41;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21110452; PubMed=11182319;  
 RA Sllakowski B., Nordstiek G., Kunze B., Blocker H., Muller R.;  
 RT "Novel features in a combined polypeptide synthase/non-ribosomal  
 RT peptidic synthetase: the myxlamid biosynthetic gene cluster of the  
 RT myxobacterium Stigmatella aurantiaca Sga15.";  
 RL Chem. Biol. 8:59-69(2001).  
 DR EMBL; AF319988; AAK57189.1; -  
 DR InterPro; IPR001227; Ac\_transferrase.  
 DR InterPro; IPR000794; Ketoacyl-synt.  
 DR InterPro; IPR006163; Pp\_bind.  
 DR Pfam; PF00698; Acyl\_transf; 1.  
 DR Pfam; PF00109; ketoacyl-synt; 1.  
 DR Pfam; PF02801; ketoacyl-synt\_C; 1.  
 DR Pfam; PF00550; Pp-binding; 1.  
 DR PROSITE; PS50075; ACP\_DOMAIN; 1.  
 DR PROSITE; PS00606; B\_KETOACYL\_SYNTHASE; 1.  
 DR Phosphotransferase; Transferrase.  
 KW Phosphotransferase; Transferrase.  
 SQ SEQUENCE 1862 AA; 200005 MM; E809DCECA320144 CRC64;

Query Match 3.6%; Score 143.5; DB 2; Length 1862;  
 Best Local Similarity 22.3%; Pred. No. 0.35;  
 Matches 180; Conservative 79; Mismatches 289; Indels 255; Gaps 37;

3 AVAAVAVVLLLAGAGAGDEAREAAVRLVAVRLGPPAPDPSVERALAAKPG 62  
 962 AVAPASALVELGSAABALGAGARELSVEPERALVLEADRRLVQVHLSPASGQHV 1021  
 63 DTVSLGGGGA-----RVVRGSGTGVAAAGLHRLRPFCCGHVAMSSQRLRP- 112  
 1022 HHSRAVGGTSPGAGVRRHCRQIRANTPSGSPVELDAVRRCAQHVGAAVYQLEPC 1081  
 113 ---PLPAAV-----PGLTEATPNRYRYQVNTOSYSFVWMDWARMERE--IDWM 157  
 1082 NVQYEAFLTLGEAMRRPGEALGVALSPEVQ-----ESARYQHPALLD-A 1128  
 158 ALNGINLALAMSGQALIMORVYLAGLQALNEPFTGPAFLAMG-----RMG 205  
 1129 GLQTLALALAAESGEAV--LPMPLSIGLECVQ--GRADVKMAHVSIAAATSPEDRMG 1182  
 206 NLHTWDG-----PLPSPWHIKQYLOH--RVLDQKRS-----FGMTPLYPAFAHVP 250  
 1183 TLELDGERRAVAVARVAVRVAARLLLEVGERBSQEMFVDVWEPRAVTAQGPARD 1242  
 251 EAVTRVFPQVNTKMGSGHF-----NCSYSCFLPAE-----DPIPIIG 292  
 1243 WLV-----FLDGGWGTALVEIGRQGCCTVVTAGETFORODARFVVPKPEDM 1294  
 293 SLEFELI-----KRGTDHYGADTFNEMQPPSEPSYLAATAVAYEAM----- 338  
 1295 ERLIRELPALPAGHBRVAVYLGDAVLDQGTGTESS--VAALHLVAKMSPPARILW 1352  
 339 ---TAVDTEAVWLLQGMFL-----QHQPFWGPAQIRAVLGAVPRGLLVLDL 383  
 1353 VYTRGAQVTVGCTERYSLAQAFLMGSGSVSLQGVWGG-----LIDL 1396  
 384 FASQPYRTTASFOGPFIMCLHFGNGHGLFGALF-----AYNGGPEAR 431  
 1397 APER--VAETAAVIRE-----ISAFGGDGEDQIALREKSLVPLIARGRVNAPAEPLR 1448

QY 432 LFNSTMGVGNAPBGISQNEVYVYSLMELGMRKDPVPLDAAWTSFAARRVGVSHPDAG 491  
 DB 1449 LPPDGLVLTGG-----LGGILGX-----VARMVARGARHL----- 1480  
 QY 492 AAMRLLRBSVNCSGEACGHRNRPVLRPSPLOMTSIWYNSDVFEAMRLLITSAPSLA 551  
 DB 1481 ---VLLGRSGASGAGAPASARRREGI-----ESLRALGASVTTTLA 1517  
 QY 552 TSPAFRYDLDLTRQAVQELVSLYEARSAYLSKELASLRAGVLAAYELLPADEVIA 611  
 DB 1518 VVVADEKXALLIREAAATL-----PRLRGVHAALLTESNLENMD--LA 1561  
 QY 612 SDSRFI---LGSWL---EQAPAAVSEADRYEONSRQTLTKPBGNIIDYANKOLAG 665  
 DB 1562 AMTAMRPRVGLSGWVLEHVTR-----EAEIDFVMSSTS-TLWGASG-LAHYA---AG 1610  
 QY 666 LVANYYTPRWRLEFLLEALV-DSVAGGIP 691  
 DB 1611 -----NOFLHALHRRAGSLP 1627

## RESULT 14

Q96W93 PRELIMINARY; PRT; 444 AA.  
 ID 096W93  
 AC 096W93  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE TR18.  
 GN TR18.  
 OS *Gibberella zeae* (Fusarium graminearum).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreales; Nectriaceae; *Gibberella*.  
 OX NCBI\_TaxID=5518;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=88-1;  
 RX MEDLINE=21318682; Pubmed=11425709;  
 RT Lee T., Oh D.W., Kim H.S., Lee J., Kim Y.H., Yun S.H., Lee Y.W.;  
 RT "Identification of deoxyribovalenol- and Nivalenol-Producing Chemotypes  
 RT of *Gibberella zeae* by Using PCR.";   
 RL Appl. Environ. Microbiol. 67:2966-2972 (2001).  
 DR EMBL; AF36365; AAKS3574.1; -  
 DR InterPro; IPR005152; LIP.  
 DR Pfam; PF03583; LIP.1  
 SQ SEQUENCE 444 AA; 47659 MW; BC5208C72484C701 CRC64;

Query Match 3.2%; Score 128; DB 3; Length 444;

Best Local Similarity 20.3%; Pred. No. 0.57;

Matches 97; Conservative 54; Mismatches 160; Indels 166; Gaps 24;

QY 157 MALNGIMLALA-MSGGEALWQRYVALGLTQAEINEFFTPAFILWGRMGNLHTWGPDP 215  
 DB 1 MALDRLIFLISLWIG-----FVGAQAALSL-----PLP 29  
 QY 216 PS--WHIKOLYLOHRLVDQMSFGMTVPVPAFAHVEAVTRVPPQV-NTVMGSGWGF 271  
 DB 30 PSKDPMTAP-----PGFENAPGVTLAVRPAFGLTSGVTG----- 65  
 QY 272 NCASVSCSFLAPEDPIFP---IGSLFLRELKEFGDHIYAGDITENQPPSEPSSTYLA 328  
 DB 66 NCASVSIILRYTTSHKPTWAVTTLLVPELGPESNNAHQKIQOSALMSLQVAYDSPDYA 125  
 QY 329 AATTAAY--BAMTAVDTEAVMLQGLFQHQPOFWGPA-----QIRAVLG 371  
 DB 126 SPSVTVYTAHISIIYEAA-LAEG-IFVSVDPYEGSNAFTAGIISGYATLDSIRAYLS 183  
 QY 372 AVPRGALLVDPRSPVYTRTASFGQGFIMQMLNFGNGLG---ALE-ANVGP 427  
 DB 184 -----LGLGFNMTNTPSVA-----LW-----GYSGAFAATEWASLAVQVAP 220  
 QY 428 EAARLFPNSTMGVGNAPBGISQNEVYVYSLMELGMRKDPVPLDAAWTSFAARRVGVSH 487

DB 221 ELVA---GPVGAAGAPLANITSIMYDNG--GPTSGFTPMILGLT-----VOY 266  
 QY 488 PDGAAMRLLRBSVNCSGEACGHRNRPVLRPSPLOMTSIWYNSDVFEAMRLLITS 547  
 DB 267 PDV-----RKHLVSK-----LNNDSQYNTDFLAAGYITFA 299  
 QY 548 PSLATSPAFRYDLDLTRQAVQELVSLYEARSAYLSKELASLRAGVLAAYELLP 604  
 DB 300 -----GVAFSKIDINK-----YFGKTDIFSDEPITALVNRBGWGYHGP 340

## RESULT 15

Q9FBV4 PRELIMINARY; PRT; 783 AA.  
 ID 09FBV4  
 AC 09FBV4  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative DNA-binding protein.  
 GN SC07844 OR SC00003 OR SC0857.41C.  
 OS *Streptomyces coelicolor*.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Brown S.P., Harris D.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; Pubmed=8843436;  
 RA Redenbach M., Kleiser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmid and a detailed genetic and physical map for  
 RT the 8 MB *Streptomyces coelicolor* A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96 (1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; Pubmed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete *Streptomyces*  
 RT *coelicolor* A3(2).";  
 RL Nature 417:141-147 (2002).  
 DR EMBL; AL939132; CAC03662.1; -  
 DR EMBL; AL939104; CAD30877.1; -  
 DR InterPro; IPR005114; Helicase\_ass.  
 DR Pfam; PF03457; HA; 6.  
 SQ Hypothetical protein; DNA-binding; Complete proteome  
 SQ SEQUENCE 783 AA; 86432 MW; 545B556395B05EDF CRC64;

Query Match 3.2%; Score 124.5; DB 16; Length 783;

Best Local Similarity 20.4%; Pred. No. 2.4;

Matches 141; Conservative 69; Mismatches 249; Indels 233; Gaps 31;

QY 4 VANAAGVVLAAAGAGAGBEAANAARLVARILGPGGAADFVSVERALAAKP--- 60  
 DB 43 LSANAAGPVGVWGRGLVVG-EVFLAAGVVRQLASLC-----QVSESHLSRALPSWD 93

```

Oy      61  -GLDTSLG-----GGGAA-----RYVVRSTGVAAAGJHRRLRP--FCGCIVAN 103
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      94  EGADARANGQPVVRPAASGSAVPAWMLGRLCTAARTGAQSAARLYSDLRSRICIRHQCV 153
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      104 SGGQLRPRLPAVPBELTEA--TPNRIRYQV-----CTOSYFV-----WMDV 147
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      154 SLDSLALDR--VTLVEGQVGLAGJPEVIRAHOSLPLLRKKSCEAAFPVACAASWMD- 211
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      148 ARWEREIDWALNGIMLALAWSGEAIPORVIALGLTOAEINEFTGPALAWGRGNL 207
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      212 VHRDEVLMPARIGRVCADLPGEVAVLAR-----DYVTEPAVAVTVLCD 258
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      208 HTWDGFLPSPMHLKOLYLRYLDDQRSGMTFVLPAPAGHVPKAVPVFPQVNVTKGS 267
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      259 RLW-----RORVLEDTN-----GQMFHTLAEV--DRL-----283
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      268 WGHFNCSYSCSFLAEDP2PFIIGSLFRLDIKESFGTDHLYGADTFENEMPSPSESYL 327
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      284 -----LTELARILGRPMIV-----EQLAASAGALF 309
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      328 AAATFAYEAMTAVDTEAVALQGMFOHQPOFQWPAQIRAV---LGAVRGSLVTLDL 383
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      310 AWPACTRRRRGAVPEEDWVAVP--VAHRPGLA-AQVREIQLDHAGNAPVGSJ-- 361
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      384 FASQPVYTTASFOGQPIFWCMLHNFGNHGLFALBAVNGGPEAAVLFPNSTMVGTGM 443
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      362 -SRLEQAVRYGLAH-----AHSYAAHHGLAV-----PKYGR--392
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      444 APEISQNEVVYSIMALEGMRKOPVDLAAMVTSFAPARRGVS-----HPDA 490
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      393 -HEFFA-----LGANL--ANQRTGYAALLPIERQAALIRIDPFMW 427
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      491 GAANRLILSRVYVNCGEACKGNHRSLVVRPBLQWNTS1WYKRSVPFPAWMLLITSAPSL 550
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      428 NGPPIISMRTY-----HRAALVHKHGLVDATAGFPQTSIATJGEM-----468
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      551 ATSHAFRYDLDLDTROAVOELSVLYEASRAYSILKEIASLIRAGCVLAVELPALDEVL 610
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      469 LHEGCSRYDLDLHVGQGRLLADGIRPAHARSAPRRKSIALAFAGGLDYAPAPAAVGHIL 528
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy      611 ASDSR-----FILGSWLEQARA-AAVASEAD 636
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      529 ATSKSTRQDGFPLGQWLMGQSRAPAAAEKXTD 560
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: February 13, 2004, 16:17:37  
Job time : 70 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: February 13, 2004, 16:14:55 ; Search time 22 Seconds  
(without alignments)  
1428.953 Million cell updates/sec

Title: US-09-836-613-2

Perfect score: 3939

Sequence: 1 MEAVAVAAVAVGLLAGAGS.....VDLAKKIFLKYPGWAGSW 743

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

328717

## ALIGNMENTS

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pdg.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pdg.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pdg.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pdg.\*  
5: /cgn2\_6/ptodata/2/1aa/PTCUTS.COMB.pdg.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pdg.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3939	100.0	743	US-09-077-354B-2	Sequence 2, Appl 1
2	123	3.1	3340	US-09-252-991A-23568	Sequence 2, Appl 1
3	121	3.1	4472	US-08-804-227C-2	Sequence 2, Appl 1
4	120	3.0	3729	US-08-804-227C-4	Sequence 2, Appl 1
5	118.5	3.0	1665	US-09-858-664A-2	Sequence 2, Appl 1
6	113.5	2.9	916	US-09-252-991A-23637	Sequence 2, Appl 1
7	113	2.9	7257	US-09-335-409-5	Sequence 5, Appl 1
8	113	2.9	7257	US-09-567-969-5	Sequence 5, Appl 1
9	113	2.9	7257	US-09-567-969-5	Sequence 5, Appl 1
10	113	2.9	7257	US-09-567-969-5	Sequence 5, Appl 1
11	113	2.9	7257	US-09-567-969-5	Sequence 5, Appl 1
12	113	2.9	7257	US-09-567-969-5	Sequence 5, Appl 1
13	113	2.9	7257	US-09-567-969-5	Sequence 5, Appl 1
14	108	2.7	1419	US-09-252-991A-31822	Sequence 3, Appl 1
15	107.5	2.7	405	US-09-252-991A-33024	Sequence 3, Appl 1
16	107.5	2.7	1621	US-08-972-927-2	Sequence 3, Appl 1
17	107.5	2.7	1622	US-08-972-927-2	Sequence 3, Appl 1
18	105	2.7	760	US-09-252-991A-18711	Sequence 3, Appl 1
19	105	2.7	1203	US-09-075-272-4	Sequence 4, Appl 1
20	105	2.7	4150	US-09-428-517-2	Sequence 4, Appl 1
21	104.5	2.7	726	US-08-313-185-49	Sequence 4, Appl 1
22	104.5	2.7	726	US-08-459-499-13	Sequence 4, Appl 1
23	104.5	2.7	726	US-08-083-614A-49	Sequence 4, Appl 1
24	104	2.6	581	US-09-221-235-8	Sequence 8, Appl 1
25	104	2.6	581	US-09-221-928-8	Sequence 8, Appl 1
26	104	2.6	581	US-09-221-527-8	Sequence 8, Appl 1
27	104	2.6	581	US-09-221-236-8	Sequence 8, Appl 1

28 104 2.6 581 3 US-09-221-416-8 Sequence 8, Appl 1

29 104 2.6 581 3 US-09-221-245-8 Sequence 8, Appl 1

30 104 2.6 581 3 US-09-163-115-8 Sequence 8, Appl 1

31 104 2.6 581 3 US-09-221-528-8 Sequence 8, Appl 1

32 104 2.6 581 3 US-09-593-553-8 Sequence 8, Appl 1

33 104 2.6 581 3 US-09-221-237-8 Sequence 8, Appl 1

34 104 2.6 668 4 US-09-617-591A-4 Sequence 4, Appl 1

35 104 2.6 677 3 US-08-836-567-2 Sequence 2, Appl 1

36 104 2.6 677 3 US-08-836-567-2 Sequence 2, Appl 1

37 104 2.6 1197 3 US-08-836-567-12 Sequence 12, Appl 1

38 104 2.6 1197 4 US-09-606-304-12 Sequence 12, Appl 1

39 101.5 2.6 2457 4 US-09-252-991A-25671 Sequence 25671, A

40 101 2.6 3519 3 US-09-428-517-4 Sequence 4, Appl 1

41 101 2.6 5215 3 US-09-105-537-2 Sequence 2, Appl 1

42 100.5 2.6 2628 3 US-09-413-814-11 Sequence 11, Appl 1

43 100 2.5 425 4 US-09-252-991A-20467 Sequence 20467, A

44 100 2.5 588 4 US-09-252-991A-31356 Sequence 31356, A

45 100 2.5 1323 1 US-08-026-138E-4 Sequence 4, Appl 1

RESULT 1

US-09-077-354B-2

Sequence 2, Application US/09077354B

Patent No. 625096

GENERAL INFORMATION:

APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;

INVENTOR: WEBER, BIRGIT; BLANCH, LINNE; ANSON, DONALD STEWART

TITLE OF INVENTION: SYNTHETIC MAMMALIAN

FIELD OF INVENTION: N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SAME

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSER: SCULLY, SCOTT MURPHY & PRESSER

STREET: 400 GARDEN CITY PLAZA

CITY: GARDEN CITY

STATE: NEW YORK

COUNTRY: UNITED STATES

ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/077,354B

FILING DATE: 22-APRIL-1999

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/00747

FILING DATE: 22-NOV-1996

ATTORNEY/AGENT INFORMATION:

NAME: POKALSKY, ANN R.

REGISTRATION NUMBER: 34,697

REFERENCE/DOCKET NUMBER: 12416

TELECOMMUNICATION INFORMATION:

TELEPHONE: 516 742 4343

TELEFAX: 516 742 4366

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 743 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

FEATURE:

NAME/KEY: Potentially-glycosylated Asn site,

LOCATION: 261

FEATURE:

NAME/KEY: Potentially-glycosylated Asn site,

LOCATION: 272

FEATURE:

NAME/KEY: Potentially-glycosylated Asn site,

LOCATION: 435

```

FEATURE: Potentially-glycosylated Asn site,
LOCATION: 503
FEATURE: Potentially-glycosylated Asn site,
LOCATION: 513
FEATURE: Potentially-glycosylated Asn site,
LOCATION: 526
NAME/KEY: Potentially-glycosylated Asn site,
LOCATION: 532
US-09-077-354B-2

```

```

Query Match 100.0%; Score 3939; DB 3; Length 743;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MEAAVAAAVGVLLLAGAGAGADEARAAAVRALVRLGPGPADPSVSVERALAAKP 60
DB 1 MEAAVAAAVGVLLLAGAGAGADEARAAAVRALVRLGPGPADPSVSVERALAAKP 60
QY 61 GLDYSLGGGGAARVVRGSGVAAAAGLHRYLRDFCGCHVAMSGSQLRPLRPVAVGE 120
DB 61 GLDYSLGGGGAARVVRGSGVAAAAGLHRYLRDFCGCHVAMSGSQLRPLRPVAVGE 120
QY 121 LREATNRRYVYONCTOSYSFYVMDMAWREIDMMALNGINILAMSGCAIWRVYL 180
DB 121 LREATNRRYVYONCTOSYSFYVMDMAWREIDMMALNGINILAMSGCAIWRVYL 180
QY 181 ALGLTOAEINEPFTGPAFLAMGSMNLHTWDGRLPPSWHIKQLYLQHRVLDQMSFGWTP 240
DB 181 ALGLTOAEINEPFTGPAFLAMGSMNLHTWDGRLPPSWHIKQLYLQHRVLDQMSFGWTP 240
QY 241 VLPAPGAVPEAVTRYFPQVNTYTKGSGWGHFNCSISCSFLAPEDPIPIISLRLRLI 300
DB 241 VLPAPGAVPEAVTRYFPQVNTYTKGSGWGHFNCSISCSFLAPEDPIPIISLRLRLI 300
QY 301 KEFGTHIYGADTFENMOPSPSEPSYLAATTAAYEAAMTAVDTEAVMLLQGLFQHQPF 360
DB 301 KEFGTHIYGADTFENMOPSPSEPSYLAATTAAYEAAMTAVDTEAVMLLQGLFQHQPF 360
QY 361 WGBAQIRAVLGAVERGRLLVLDLFAESQPVYTRTASFOGQPIWCMLEHNFSGNHLFGAL 420
DB 361 WGBAQIRAVLGAVERGRLLVLDLFAESQPVYTRTASFOGQPIWCMLEHNFSGNHLFGAL 420
QY 421 EAVNGGPEAARLFPNSTVMGTGMAPRGISQNEVYISLMAELGMRKQDPVDLAAMVTSFA 480
DB 421 EAVNGGPEAARLFPNSTVMGTGMAPRGISQNEVYISLMAELGMRKQDPVDLAAMVTSFA 480
QY 481 RRYGVSHPDGAAWRLLRVYVNCSEACRGNHRSPLVRPSLQNTSITWYRSDVFEAM 540
DB 481 RRYGVSHPDGAAWRLLRVYVNCSEACRGNHRSPLVRPSLQNTSITWYRSDVFEAM 540
QY 541 RLLLTAPSLATSPARVYLDLITRQAVQELVSLYEERSAVLSKELASLRAQGYLAX 600
DB 541 RLLLTAPSLATSPARVYLDLITRQAVQELVSLYEERSAVLSKELASLRAQGYLAX 600
QY 601 ELLPALDEVLASDSRFLGSMLEQARAASVSEADDFEONSRVYQTLTGPGENITLDYAN 660
DB 601 ELLPALDEVLASDSRFLGSMLEQARAASVSEADDFEONSRVYQTLTGPGENITLDYAN 660
QY 661 KQLAGIVANYTYTRKMLFEALVDSYAGQIPFOQHOFPDKNVQOLBQAFVLSQKRPSPQR 720
DB 661 KQLAGIVANYTYTRKMLFEALVDSYAGQIPFOQHOFPDKNVQOLBQAFVLSQKRPSPQR 720
QY 721 GDTVDLAKKIFLKYYPGWYAGSN 743
DB 721 GDTVDLAKKIFLKYYPGWYAGSN 743

```

RESULT 2  
US-09-252-991A-23568

```

Sequence 23568, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 23568
LENGTH: 3340
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23568

```

```

Query Match 3.1%; Score 123; DB 4; Length 3340;
Best Local Similarity 21.4%; Pred. No. 0.11;
Matches 178; Conservative 10; Mismatches 318; Indels 236; Gaps 44;

```

```

QY 4 VAVAA-----AVGVLLLAGAGA-----AGDEAREAAVRA-LVARTL 40
DB 1048 VAIERSPOLLVGLLAIVAGGAYVLPDYPSERLAYVLADSGVELLITQAHLEPRL- 1106
QY 41 GPG-----PADPSVSVERALAAKPGLDYSLGGGGAARVVRGSGVAAAAGLHRYLRD 95
DB 1107 -PGAGVTPICLDLSLKIDNMPSOAPGLHLH--GDNLAYVYTSGSISQPKVG----- 1156
QY 96 FCCGHVAMSGSQLRPLRPVAVGE--LREATNRRYVYONCTOSYSFYVMDMAWRE 153
DB 1157 --NTHALLA-ERLQNMATATALDGDVLMQAP-----VSFDVSW--E 1195
QY 154 IDMMALNGINILAMSGQ-----EAIWRVYLAIGLTOAEINEPFTGPAFLAMGRMG 205
DB 1196 CFMFLVYGCGLVLAAPGHHDPARLVELVROFVTTLHFPVPLLOFIDEPFGAAGSILR 1255
QY 206 NLHTWDGRLPPSWHIKQLYLQHRVLDQMSFGWTPVLPAPGAVPEAVTRYFPQVNTYTKM 265
DB 1256 RLFSGGALPAA-----LRNRYLQR-----LPAAHLNRGPTET--AINVT-- 1295
QY 266 GSWGHFNCSISCSFLAPED---PIFPIISLRLLEKFGTDHIYGADTFENM--- 317
DB 1296 ---HWGC-----RAEDGERSPIORPLGNVLCVLDLAEF--NILPAGVAGELCIGG 1340
QY 318 -----QPPSSEPSYLAATTAAYEAMTAVDTEAVMLLQGLFQHQPFQWFG- 362
DB 1341 LGLARGYLGRPALSAERFVADPLSAAGERLYRTGDRAMNADVLEVLGRLDQVLRGE 1400
QY 363 ---PAQIRAVLGAVERGRLLVLDLFAESQPV--YTRTASFOGQPIWCMLEHNFSGN 414
DB 1401 RIEPEIQLALLQPGVAAQVAVVIREGVAGSOLVGYTTGAVGAEEA-----EQNQ 1451
QY 415 GLFGALEAVNGGPEAARLFPNSTVMGTGMAPRGISQNEVYISLMAELGMRK---DPVPD 470
DB 1452 RLRAALQA-----ELPEYVWPAQMLRLAOWPLPS--GKLDTRALPEVWMOQRHVEPRT 1505
QY 471 L-----AAMVTSFAARYG-----VSHPDGAAWRLLRVYVNC-- 505
DB 1506 LQRIANVINSBVGLPRVGLRDPFELGHSLLATRIVSTRQACVDELPLRLAFESSEL 1565
QY 506 -----GEACRGNHRSPLVRPSLQNTSITWYRSDVFEAMRLILTSAPSLATSPAP 566
DB 1566 EAPCEQVRAAQAAGRTDSHAIIRIDRECPVPSYSQQRWFMFMOL-----EPD---SPAY 1618
QY 567 RY-----DLIDLTR--QAVQELVSLYEERSAVLSKELASLRLA---GGV-LAYELL 603
DB 1619 NVGGLARLSGPLDARFEALLQALVQ--RHETLRTTPPSVDVGVQKRVHGGGLHMDKDF 1677
QY 604 PALD-----EVLASDSRFLGSMLEQA---RAAAVSEADDFEONSRVYQTLTWGPE 652

```

DB 1678 SALDRSROQHLOTLA-DSEARHPDLESGLPLRVCMKARERHYLVTLHIV---TE 1733  
QY 653 GNLDYANKOLAGVANYTYPRMFLFLEALVDVAQGFPOQHOFDQNVQLE 705  
DB 1734 GWANDIPARELGALYEAFLDLR-----ESPLSELPVO-YLDYSVMORE 1775

## RESULT 3

US-08-804-227C-2  
Sequence 2, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII (DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804.227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4472 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-804-227C-2

Query Match 3.1%; Score 121; DB 2; Length 4472;  
Best Local Similarity 21.6%; Pred. No. 0.29;  
Matches 176; Conservative 71; Mismatches 280; Indels 288; Gaps 44;  
QY 4 VAVAAV---GVLILAGAGAGADDEARAAVAVLVARLLG-----PGRAAD 47  
DB 1686 VDVAAVVHSGGELIAAAVAGALSLD-DAAAVVALRAGLIGRYLAGAMAAVPLPGE 1743  
QY 48 FSVSVERALAAKPGIDTSLGGGAA-----RVVRSGSTVAAAGLH----- 90  
DB 1744 -----VEAGLAKPGVEVAAVAVGPASTVSGRRVAVAGVAVCAEGVQALIPVDYASHS 1799  
QY 91 RYLRDFCG-CHVAMSGQLRPR-PLPRAV-----PGELTEATPNRRYRYQVNTQSTYSFV 143  
DB 1800 RHVEDLKKELEKRVLSGIRSPRPVVCSTVAGEOPGEFV----- 1838  
QY 144 WMDVARMER-----EIDWMLNGINL-----ALAMSGEAIWQRYVLA 181  
DB 1839 -FDAGYVRNIRNRYEBAVAVGGLBGRKRFIVSAHPVVALEQTAELAA--DRSVHA 1895  
QY 182 LGLTQAE---INEFFTGAFLANGMGMLATWDGPLPSPWHIKQLYL-----QHRVIDQM 233  
DB 1896 TGTARRQDSDPHRLITSTA-EAMAHGATL-TWDBALPFG-HLTTLPTYPFNHHHYWMD-- 1950

QY 234 RSEGMTVPLPAPAGHPEAVTVFPQVNTYKKGSGHFNCSYSGFLAPEDPIFP-ITG 292  
DB 1951 -----TTPPTPATTTOSPTDAMR-----YRVT-----WK-----ALTSSSPVRPHSIG 1988  
QY 293 SLFLRELKEPGTDHVIYGADTFENOMQPSSEPSVLAATAYEAMTV-----D 342  
DB 1989 RCLL-----VAPPTDDELDDGLTTVISEGASVARELEVIGARR 2028  
QY 343 TEAVMLQGLMFOHOPQFWMGAQ-----IRAVLAGVPRGLLVLDPAESQPYRTASFO 398  
DB 2029 AEVVAL-----KPSMSAGEENTTVYSLGLVP-----STDVARTSIALIQ 2070  
QY 399 -----GQPF--IWCL-----HNFQNGHGLGALAAV-----GG-----PEAA 430  
DB 2071 AVSDIGVPAARVWALTTRRAVAVPGETPDQAGQLMGFRVALLLPDIMGLDILPETA 2130  
QY 431 RLEPNSTWGTGMADEGISQ--NEVYSIMAEIGRKDPVPDIAAMVTSFAARVGVSHP 488  
DB 2131 ELTRTFETSQPPQTERLPQTNRRALBELAAVLAGRGEQDVAVRASGIYGRVSRPAA 2190  
QY 489 DAGAAMR---LLASVYNGSGEACRGHNSPLVRPSLQNTSIVNRSDFEAMRLIL 544  
DB 2191 AGAASWQPSGVLITGMAIGR-----IARRLAAE-----GAERLVL 2229  
QY 545 TS-----APSLATSPAFRYDLD-----LTRQAOELVSLY----- 575  
DB 2230 TSERGEPAPGA-----ELAEIRGHGCEVYHAACVABDALAAVTAVPRAVHT 2282  
QY 576 ---YEBARSAYLSKELIASLRAGVLAAYELLPALDEVLASDSRFLI-----GSWEQARA 627  
DB 2283 AGILDVAVIDTSPSFEVTRGAKVCAEELLHQLTADIKGLDAFVLFSSVTGTWGNAGOG 2342  
QY 628 A-AVSEAEADPYEONS-----YQLTMGPEG 653  
DB 2343 AYAANAALDUALERRRAAGLPATSVAMGLMGGG 2377

## RESULT 4

US-08-804-227C-4  
Sequence 4, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII (DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804.227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3729 amino acids

TYPE: amino acid  
 TOPOLOGY: unknown  
 MOLECULE TYPE: protein  
 US-08-804-227C-4

Query Match 3.0%; Score 120; DB 2; Length 3729;  
 Best Local Similarity 22.4%; Pred. No. 0.26;  
 Matches 192; Conservative 72; Mismatches 265; Indels 328; Gaps 47;

4 VAVAAV-----GYLLAGAGAGADBARAAAVRALVRLG-----PGPAD 47  
 668 VDVAAVVGSGGIIAAATVAGLSLE--DAAAVLALRIGRIYLAGGMAAVLPAGE 725  
 48 FSVSVERALAAKGLDLYSLGGGAA----RVVVGSTGVAAAGLH-----90  
 726 ---VEAGLAKMGVAVAAVNGPASTVSGDRRAVAGVAVCAAGVCAALLPYDVASHS 781  
 91 RYLDFCG-CHVMSGSQLRPR-PLPAV-----PGLTEATPNRYRYQVCTQSYSFV 143  
 782 RHVEDLKGELEKRLSGIRPSRPVPCSTVAGEQPEPV-----820  
 144 WMDMAWNER-----EIDMMLNGINL-----ALMSGQEAIWQRYVLA 181  
 821 -FDAGYWFNLRVRVFFSAVVGILBEGRRFTEVSAHPVLVAIEQTAEAA--DRSYHA 877  
 182 LGITQAE---INRFTGPFLAMGRMGNLHTWDGRLPSPWHIKOYL-----QHRVLDQM 233  
 878 TGLRQODDSPHRLITSTA-EAAMAGATL-TWDPALPFG-FLTTLPYTFNHHWLD--932  
 234 RSFGMPFVLPFAF-----AGHVPEAVT--RVFPQVNTKM--GSMGFNCSYCS 278  
 933 ---TPTTPATTTQSPDAQNPAADALPKVSMKRLRDDSLTARLDGM-----978  
 279 FLIAPEDFPIPTIGSLFLAEELIK-----EFSTH-----IYGADTFNEMQPS 322  
 979 LTVVPPASADPSVAGVARELTARGATVESLTVEPPADSRRLGILLVDAETERDEAGPIRG 1038  
 323 EPEYLA-----AATVAYEAMTAVDTEA-VMLLQGMFLQHPQFPG 362  
 1039 IVSLLLAGDHAGADGARPVPAGLAASLALIQADAGTEAGLM-----1083  
 363 PAQIRAVLGAVPFGRILLVDLFAESQPV--YRTTASFGQPFITGMHLNFGNHLGFG-418  
 1084 -AVTRGAVVAVPG-----DVPAPSQALLMGFRVAGIG-LPHCWGLDLDPFGDSDGF 1135  
 419 ---ALEAVNGGPEAARLFPNSTWGT-GMAPEGISQNEVYSIMAEI 461  
 1136 RQLAATLAGRAPADQVALRA--SGAVGRLVRAASAGAGADGWRPRG-----TVLV 1184  
 462 GMRKDPVPLAAVWTSPAAARY---GVSHP---DGAAMRLILRSVNGSGEACRGHNR 514  
 1185 GDTAEVAGPLVFWLNGARRVTLISGLSGPLPEELADVAVRVTV-----APCDPADR 1236  
 515 SPLVRPSLQMTNISIWNISDVFEAMRLLTSPASLATSAPARYDLDTLROA-----567  
 1237 PALRTLLAABAPTAV-----LVAPPVPTP-----LAEMTAALALIALISA 1277  
 568 ---VOELVSLYE-----EARSAYLSKELASIL 592  
 1278 KTGLVRLDLSLDEPPLLEDGLDAFVVFSSVAVGWAGGQGYAAGTAYLD-ALAEGR 1336  
 593 RAAG--VLAVELLPADEVLTASRFLGSKLEQA-----RAAAV 630  
 1337 RAAGLPTVSAAWTFWIG-TPADS---LGGQSRAGITPLDPAASIDALARAVERRAGCV 1392  
 631 SEAEADFYEQNSRYQLT 647  
 1393 TVADIDMERFASAYTAT 1409

RESULT 5  
 US-09-858-664A-2  
 ; Sequence 2, Application US/09858664A

Patent No. 6482624  
 GENERAL INFORMATION:  
 APPLICANT: WEI, Ming-Hui, et al.  
 TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
 TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
 TITLE OF INVENTION: THEREOF  
 FILE REFERENCE: C1000927-C1P  
 CURRENT APPLICATION NUMBER: US/09/858,664A  
 PRIOR APPLICATION NUMBER: 2001-05-17  
 PRIOR FILING DATE: 2000-11-11  
 NUMBER OF SEQ ID NOS: 33  
 SOFTWARE: PASCSEQ for Windows Version 4.0  
 SEQ ID NO 2  
 LENGTH: 1665  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-858-664A-2

Query Match 3.0%; Score 118.5; DB 4; Length 1665;  
 Best Local Similarity 18.6%; Pred. No. 0.094;  
 Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

19 GGAAGDARAAVRAVRLVRL-LGPGPAD---FSVSVERALAAKGLD-----TY 65  
 815 GSASQASSQVSSLRVSSQVTEBPSPSLAEGMTGEADLSDSTPLQRPOCVTWKRF 874  
 66 SLGG-GGAARVVRGSGTVAAAGLHRYLDFCG-----CHVMSGSQLR-----109  
 875 SLGGGGVAGVAGVGTAFPGDAG-----GMLGGQMMARITAAVSSQSEEBQEARAE 928  
 110 ---LPLPAVPELTPNRYRYQVCTQSYSFVWMDMAWNE 151  
 929 SQSEEQEARAABSLPVSARPIPEVGRAPTSS-----EPTPWE 969  
 152 ---REI--DMMLNGINLAL-----AMSGQEAIWQRYVLA 182  
 970 DIGQSVLQIRDSGDLEADDTISLDSVDPAYLNLSDYDIKYLPFEEMIRKPKSA 1029  
 183 ---GLTQAEINEF-----FTGP-----AFLAMGSMGNLHTWGP 213  
 1030 QPEPSPMAEBELAEFPPTWPMNGELGPHAGLEITBESDVALLAEAAVGRKRWSSP 1089  
 214 LPSPWHIKQDYLQHRVLDQNRSGMTFVLPAPAGHV-----PEAVTRVFPQVNTKMG 266  
 1090 SRSLEHPPG---RHLPLDEBAELGRERYKASVEHISRLIKGRPEGLEKGP---RKRP 1143  
 267 SMGFNCSYCSFLLAEDDIPITIGSLFLRELKERTDHTIGADTFNEMQPSSEPSY 326  
 1144 GLASFRLSGLKSMDRAP-----TFREL-----SDETVV 1172  
 327 LAAATVAYEAMTAVDTEAVWLQGMFLQHPQFPGAPQIRAVLGAVPRG--RLVLVDLF 384  
 1173 LGQSVTLACQVSAQPAQATWSKDGAPLE-----SSSRVLTISATLKNFOLLTLIVV 1224  
 385 AESQPYTTRTNSFGQPFITGMHLNFGNHLFGALEAVNGGEAARLFPNSTWGTGMA 444  
 1225 AEDLGVT-----CSVSNALGVTYTTGVLR-----KAER--BSSS 1257  
 445 PEGISQNEVVYSIMAEIGMRKDPVPD-----LAAW--VTSFAARYGVSHPDAGAA 493  
 1258 ---WTLASDIFDCCYLTSLKSGGYTFTTACVSKAGMPYSSPSEQVLLGSPSHLASEES 1355  
 494 WRLLRSVYNC-----PCPDIGEVYADGVLLVMKPVSSYGVTVIVOCSLGGGS 1295  
 1296 WTLASDIFDCCYLTSLKSGGYTFTTACVSKAGMPYSSPSEQVLLGSPSHLASEES 1355  
 510 RGHNRSPVRRPSLQMTNISIWYNR--SDVFAARLILTSAPSLATSAPARYDLDTLR--Q 566  
 1356 QGRSAQPLPSTKTPATPOTQIRGFSVYRQCEWKAASRALAAKIPHPDKXAVLREYE 1415  
 567 AVQELVSLYYEARSAYLS--KELASLIRAGGVLAIELLPALDEVLTADSFLGSKWLEQA 625

Db 1416 ALKGRLHPLAQLHAAYISPHLITLLEL--CGDELLPCLAE--PASYSESEVDYIMQM 1472  
 QY 626 AAAA-----VSEADAEYEQN---SRVQTLTGEGNITLDYANKQLAGLVANYYTPMRL 677  
 Db 1473 LSATQYLHNOHLHLDLSENNIITEVYLL-----KVYDIGNQSLSEKVLPSDKFKD 1566  
 QY 678 FLEALVDSVAGQ---IPFQCHOFKXNQQL--EQAFVLSKQRYEQCPFG 721  
 Db 1527 YLETPAPBLLEGGQAVP-----QTDIWAIGVATFIMLSAEYVSSSG 1568

## RESULT 6

US-09-252-991A-23637  
 ; Sequence 23637, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.116  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 23637  
 ; LENGTH: 916  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-23637

Query Match 2.9%; Score 113.5; DB 4; Length 916;  
 Best Local Similarity 21.4%; Pred. No. 0.11;  
 Matches 167; Conservative 77; Mismatches 247; Indels 289; Gaps 42;

QY 1 MEAVVAAAVGVLLLAGAGAA-----GDEAREAAV---RALVAR-----LL 40  
 Db 238 LQARLAAALGA--GEGAGVAAEQAFQVLRQRRARQRRVPTQCGAAARHOFLA 295  
 QY 41 GPGPAFDSVVERALAKPGLDTYSLGGGAAR-----VVRGSGTGAATA 87  
 Db 296 GTGLAVDQRRVERHRLARTGLQ-----GADRLGIAEHLBPLGVMVQGRQALDAV 348  
 QY 88 GLHRLRDFCGCHVAMSGSO-----LRLPRPLPAVPELTEATPNRYRYOVCTQSYS 141  
 Db 349 ---RFV-----EGQAAAGVFLALARPFGDRGIEQQGLAHGHPQR----- 388  
 QY 142 FVWMDMAREREIDMALNGINLALAWSGQBAIWQVYLAIGLQAEINEFTGPAFLAW 201  
 Db 389 ---QAELLDRIIGFEIVEQA--GQPLAGR-----VASALAE 422  
 QY 202 GRMGLTLHTWDBPLRPSWHIKOLYLQHRVLDQMSFGMTFVLPAPAGHVPNAVTRVFQVN 261  
 Db 423 GRVGOH-----QALVAHROHRI-----GHRQQRLELQAAAL 455  
 QY 262 VTKMGSMGH---FVCSYSCFLLAPEDPIPIIGSLFLRELIKEFGTD----- 306  
 Db 456 AGKVDHHRHLHATDAQORCA-----QFLHFGEGRRVDVVRH 496  
 QY 307 ---HITYG--ADTFENEMQPPSSBSYIAAATTVTEAMNAVTEAIVML 349  
 Db 497 HFRHIOVAVAPLQORHLLGDADTVDEGDVAHAGTW--TAATILAIKESIVAPGA-- 550  
 QY 350 QGMFLFOHOPFWGPAQO---IRAVLGAVPRGLVLDLFAESOPV--YTRTASFGQGFIV 404  
 Db 551 ---QPK--GRPSGGAQGLAAGAVAPALVHHGFGHQRARLLRVGLVQ----- 595  
 QY 405 CMLHNFGNHGLFGALFAVNGGPEARLLFPNSTVY-----GTGAAPGEGISQNEVYS 456  
 Db 596 ---HFGG--LGVVPGGGD--VRQHPRHLAVADPQGRRLPGAGPPTQQLQALAA 643

QY 457 LMAE-----LGRKDE--VPDLAMVTSFAAR-----RYG 484  
 Db 644 AVEAEADEBHRAGATAQRLG--RDARVGAPEVAAILFVVAQCGAVRRVGHYPGEVRG 701  
 QY 485 VSHPDGAAMRLLRLRESVNCSGEGACHNRSPVLRPSL--QMNSTWNRSDVPEAWML 542  
 Db 702 TG--EHAASQRYAR--VDGAGDVALADBAFGIARPGALGQBOPRITVQAS--VEAY-- 754  
 QY 543 LITSAPSLATSPAIFYD-----LLDT-----TROAVOELVSLYEARSAYLS 585  
 Db 755 LIDEALHVAFAVRAHDLRRSDALLEVGMAGBEGLLGGEGALRLRALIVQARRIDPF 814  
 QY 586 KELASLLRAGVLAAYLLPALDEVLASDSRFLLGSLWLEBARAAVSEADAEYEQNSRQ 645  
 Db 815 QGLAALPVGG-----LPLAQ--AED--FAFGSLDQ--RVQVQEEGAEYFAQSAHR 863

## RESULT 7

US-09-335-409-5  
 ; Sequence 5, Application US/09335409  
 ; Patent No. 6121029  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schupp, Thomas  
 ; APPLICANT: Ligon, James  
 ; APPLICANT: Molnar, Istvan  
 ; APPLICANT: Zirkle, Rose  
 ; APPLICANT: Cyr, Devon  
 ; APPLICANT: Goetlich, Joern  
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLIONES  
 ; FILE REFERENCE: 4-30582A  
 ; CURRENT APPLICATION NUMBER: US/09/335,409  
 ; CURRENT FILING DATE: 1999-06-17  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 5  
 ; LENGTH: 7257  
 ; TYPE: PRT  
 ; ORGANISM: Sorangium cellulosum  
 US-09-335-409-5

Query Match 2.9%; Score 113; DB 3; Length 7257;  
 Best Local Similarity 22.3%; Pred. No. 4;  
 Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPSWHIKOLYLQHRVLDQMSFGMTFVLPAPAGHVPNAVTRVFPQVNTKX--GSGWHN 272  
 Db 4805 IPAEMPLQGVITHAAGALD-----GVLDQTTDRFGRVLA PKVTGAW----- 4846  
 QY 273 CSYSCGFLLAPEDPIPIIGSLFLRELIKEFGTDHIYGADTFENEMQPPSSBSYIAAAT 332  
 Db 4847 ---NLHETLAGNDLAFVFLFS--SMGGLLSAGQSNVAAANTFLDLAAHRRAREGLAAGSL 4902  
 QY 333 A-----VYEAMNAVDTAEAVMLQGMFLFOHOPFWGPAQIRAVLG--AVPRGL--L 379  
 Db 4903 AMGPMSDGMMAAGLSA-----LDARLARHGMGLSAPQGTALLGQALAPETQLGM 4955  
 QY 380 VIDLFAESQPVYTRTASFGQGFIVCMLNFGNGHGLFGALDAVNGGPEARLLFPNSTVY 439  
 Db 4956 SLDVRAAQ-----ASGAAPVFWRALVRABARHTAQAQAL-----AARL----- 4997  
 QY 440 GTGMAPEGISQNEVYSLMAE-----LGRM-----KQVPPL-----AAVTSFAAR 482  
 Db 4998 --GALPEARADBEVRKVQAEIARVLSMSASAVPVDRSLDGLDLSLTAVELRVNLQOR 5055  
 QY 483 YGVV-----HPDGAAMRLLRLSVNCSGEGACHNRSPV-----RRP 521  
 Db 5056 VGATLTPATLAPHPTVDALTRWLDKVLAVAPSVSAKSPQVALDERIALIIGICRPP 5115  
 QY 522 SLQMNSTWNRSDVPEARLLLTGAPSLATSPARVYLLDITRAVQDELVLVE--E 578  
 Db 5116 G-----GVADEPSFWRLLESGSDAVEVHERWD-----IDAFYDPPD 5154  
 QY 579 ARSAYLSKELASLLRAGVLA-----YELLPALDEVLASDSRFLT--GSMLEQARAA 628

Db 5155 VAGKMT-----RGGFLSDIRDPAPFGISPREATTMDPQQRLLLETSWEAFERAG 5207  
QY 629 AVSE 632  
Db 5208 ILPE 5211

RESULT 8  
US-09-568-102-5  
; Sequence 5, Application US/09568102  
; Patent No. 6346404  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goelach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,102  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 7257  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-102-5

Query Match 2.9%; Score 113; DB 4; Length 7257;  
Best Local Similarity 22.3%; Pred. No. 4;  
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPSPMHKOLYQHRVLDQKRSFGMTPLPAFAGHVEAVTVFPOVAVTKM-GSWGHN 272  
Db 4805 IPAMPLOQVITHAAGLDD-----GVDEQTTDFSRRLAKVTGAM----- 4846

QY 273 CSYSCFLAPEDPFIPIIGSLFRLIKEFGTDHYGADTENMOPPSSEPSYLAATT 332  
Db 4847 ---NLHETLGNLDLAFVLES-SMSGILSGAGOSNYAAANTFLDALAARRAAGLAAQSL 4902

QY 333 A-----VEAMTAVDTAEVWLQGMLEFQHPQFPGPAQIRAVIG---AVPRGL--L 379  
Db 4903 AWGFWSDGGMAAGLSA-----LQRLARHGAGLSPAQGTALLGQALARPETQLGAM 4955

QY 380 VLDFEASQPYVYRTASFOGQPIWCMLEHFGNGHGLFGALBAVNGGPEARLFPNSTV 439  
Db 4956 SLDVRAASQ-----ASGAAPVPVWRALVRAEARHTAAGQAGL-----AARL----- 4997

QY 440 GTGMAPEGISQNEVYSLMAE-----LQWR-----KDPVPDL-----AAWTSFAARR 482  
Db 4998 --GALPEARRADEVRKVVQAEIARVLSWSAASAVPVDRLSDGLDSTLTAVELRNVLGQR 5055

QY 483 YGVS-----HPDAGAAWRLLLRSYVNGSGEACRGHNSPLV-----RRP 521  
Db 5056 VGATLPATLAFDHTVTLTRWLLDKYLAVERSVSASXSPVALDEPIAIIIGICRFP 5115

QY 522 SLQWNTSIWYRSDVFEAMRLLITSAPSLATSPAFRDLDLTRQAVQELVSLYE---E 578  
Db 5116 G-----GVADPESFWRLLEEGSDAVVEVPEHERWD-----IDAFYDDPD 5154

QY 579 ARSAYLSKELASLLRAGVLA-----YELLPALDEVLASDSRFL--GSMLEQARAA 628  
Db 5155 VRGKMT-----RFGGFLSDIRDPAPFGISPREATTMDPQQRLLLETSWEAFERAG 5207

QY 629 AVSE 632  
Db 5208 ILPE 5211

RESULT 9  
US-09-567-969-5  
; Sequence 5, Application US/09567969  
; Patent No. 635457  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goelach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/567,969  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 7257  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-567-969-5

Query Match 2.9%; Score 113; DB 4; Length 7257;  
Best Local Similarity 22.3%; Pred. No. 4;  
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPSPMHKOLYQHRVLDQKRSFGMTPLPAFAGHVEAVTVFPOVAVTKM-GSWGHN 272  
Db 4805 IPAMPLOQVITHAAGLDD-----GVDEQTTDFSRRLAKVTGAM----- 4846

QY 273 CSYSCFLAPEDPFIPIIGSLFRLIKEFGTDHYGADTENMOPPSSEPSYLAATT 332  
Db 4847 ---NLHETLGNLDLAFVLES-SMSGILSGAGOSNYAAANTFLDALAARRAAGLAAQSL 4902

QY 333 A-----VEAMTAVDTAEVWLQGMLEFQHPQFPGPAQIRAVIG---AVPRGL--L 379  
Db 4903 AWGFWSDGGMAAGLSA-----LQRLARHGAGLSPAQGTALLGQALARPETQLGAM 4955

QY 380 VLDFEASQPYVYRTASFOGQPIWCMLEHFGNGHGLFGALBAVNGGPEARLFPNSTV 439  
Db 4956 SLDVRAASQ-----ASGAAPVPVWRALVRAEARHTAAGQAGL-----AARL----- 4997

QY 440 GTGMAPEGISQNEVYSLMAE-----LQWR-----KDPVPDL-----AAWTSFAARR 482  
Db 4998 --GALPEARRADEVRKVVQAEIARVLSWSAASAVPVDRLSDGLDSTLTAVELRNVLGQR 5055

QY 483 YGVS-----HPDAGAAWRLLLRSYVNGSGEACRGHNSPLV-----RRP 521  
Db 5056 VGATLPATLAFDHTVTLTRWLLDKYLAVERSVSASXSPVALDEPIAIIIGICRFP 5115

QY 522 SLQWNTSIWYRSDVFEAMRLLITSAPSLATSPAFRDLDLTRQAVQELVSLYE---E 578  
Db 5116 G-----GVADPESFWRLLEEGSDAVVEVPEHERWD-----IDAFYDDPD 5154

QY 579 ARSAYLSKELASLLRAGVLA-----YELLPALDEVLASDSRFL--GSMLEQARAA 628  
Db 5155 VRGKMT-----RFGGFLSDIRDPAPFGISPREATTMDPQQRLLLETSWEAFERAG 5207

QY 629 AVSE 632  
Db 5208 ILPE 5211

RESULT 10  
US-09-568-480-5  
; Sequence 5, Application US/09568480  
; Patent No. 6355458  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas



```

; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-568-472-5
Query Match      2.9%; Score 113; DB 4; Length 7257;
Best Local Similarity 22.3%; Pred. No. 4;
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPPSMHIKQLYGHRVLDQKRSFGMTFVLPAPAGHVEATRVPPQVNTKM-GSMGHFN 272
DB IPAEWPLQGVTHAAGLDD-----GVLEQTTDFSRVLAKVTGAM----- 4846
QY 273 CSVSCFLLAPEDPIPIIGSLFLRELKEFGTDHLYGADTFNEMQPPSSEPSYLAATT 332
DB 4847 ---NLHETLGNLAFVLFSS-SMSGILGSAQGSNYAANTFLDLAAHRAAGLAAQSL 4902
QY 333 A-----YEAMTAVDTFAVWLQGMFLFOHOPQFPGPQIATVIG---AVPRGRL--L 379
DB 4903 AMGPWSDGMAAGLSAA-----LQRLARHGMGALSPAQGTALIGQALARPETQLGAM 4955
QY 380 VLIDFAESQPYRTTASFOGQPIWCMLNHFGNGHGLFGALEAVNGGPEARLFPNSTWV 439
DB 4956 SLIDVRASQ-----ASGAVPFWWRALVBAERHTAAGQGL-----AARL----- 4997
QY 440 GTGMAPEGISONEVYSLMAE---LGWR-----KDPVPDL-----AAVTSFAAR 482
DB 4998 --GALPEARADEVRKVVQAEIARVLSWSAASAVPVDRLSDGLSLTAVELRVLGQR 5055
QY 483 YGVS-----HPDGAARLLRSVYNGSGEACGHNRSPLY-----RRP 521
DB 5056 VGATLPATLAFDHTVDALTRWLLDKVLAAPESVSASKSPVALDEPIAIGICRFP 5115
QY 522 SLQNTSITWNRSDVEAMRLLTSAPSLATSPAFRYDLIDLTRQAVQELVSLYE---E 578
DB 5116 G-----GVADPESFWRLLEEGSDAVVEVHERMD-----IDAFYDPPD 5154
QY 579 ARSAYLSKELASLLRAGVLA-----YELLPALDEVLASDRFL--GSMLEQARAA 628
DB 5155 VRGKMTT-----RFGGLSDIDRFPDPAFGISPREATTMPDQRLLETSMWAEFRAG 5207
QY 629 AVSE 632
DB 5208 ILPE 5211

RESULT 13
US-09-567-899-5
; Sequence 5, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Lagon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-567-899-5
Query Match      2.9%; Score 113; DB 4; Length 7257;
Best Local Similarity 22.3%; Pred. No. 4;
Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

```

```

Matches 108; Conservative 51; Mismatches 183; Indels 142; Gaps 22;

QY 214 LPPSMHIKQLYGHRVLDQKRSFGMTFVLPAPAGHVEATRVPPQVNTKM-GSMGHFN 272
DB IPAEWPLQGVTHAAGLDD-----GVLEQTTDFSRVLAKVTGAM----- 4846
QY 273 CSVSCFLLAPEDPIPIIGSLFLRELKEFGTDHLYGADTFNEMQPPSSEPSYLAATT 332
DB 4847 ---NLHETLGNLAFVLFSS-SMSGILGSAQGSNYAANTFLDLAAHRAAGLAAQSL 4902
QY 333 A-----YEAMTAVDTFAVWLQGMFLFOHOPQFPGPQIATVIG---AVPRGRL--L 379
DB 4903 AMGPWSDGMAAGLSAA-----LQRLARHGMGALSPAQGTALIGQALARPETQLGAM 4955
QY 380 VLIDFAESQPYRTTASFOGQPIWCMLNHFGNGHGLFGALEAVNGGPEARLFPNSTWV 439
DB 4956 SLIDVRASQ-----ASGAVPFWWRALVBAERHTAAGQGL-----AARL----- 4997
QY 440 GTGMAPEGISONEVYSLMAE---LGWR-----KDPVPDL-----AAVTSFAAR 482
DB 4998 --GALPEARADEVRKVVQAEIARVLSWSAASAVPVDRLSDGLSLTAVELRVLGQR 5055
QY 483 YGVS-----HPDGAARLLRSVYNGSGEACGHNRSPLY-----RRP 521
DB 5056 VGATLPATLAFDHTVDALTRWLLDKVLAAPESVSASKSPVALDEPIAIGICRFP 5115
QY 522 SLQNTSITWNRSDVEAMRLLTSAPSLATSPAFRYDLIDLTRQAVQELVSLYE---E 578
DB 5116 G-----GVADPESFWRLLEEGSDAVVEVHERMD-----IDAFYDPPD 5154
QY 579 ARSAYLSKELASLLRAGVLA-----YELLPALDEVLASDRFL--GSMLEQARAA 628
DB 5155 VRGKMTT-----RFGGLSDIDRFPDPAFGISPREATTMPDQRLLETSMWAEFRAG 5207
QY 629 AVSE 632
DB 5208 ILPE 5211

RESULT 14
US-09-252-991A-31822
; Sequence 31822, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31822
; LENGTH: 1419
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-31822
Query Match      2.7%; Score 108; DB 4; Length 1419;
Best Local Similarity 21.9%; Pred. No. 0.76; Indels 168; Gaps 23;
Matches 100; Conservative 40; Mismatches 148;

QY 349 LQGMFLFOHOPQFPGPQIATVIGAVP-----RGLVLIDFAESQPYRTTA---SFOG 399
DB 178 LAGMARBAQVLRG---LRIGIGAPRGEMWDGNNLLIDTPRTDAQVEALGERLQG 234
QY 400 ---OPTWCMLNHFGNGHGLFGALEAVNGG-----PEARLTPNSNMGTGMAPE 446
DB 235 GTQPMI-----RVGGGPGFGLNLPALGGLVGYVYANGGANLRLLFEAVQRMHAGLPVD 289

```

```
QY 447 GISONVVYSIMAEIGWRKDPV-----DLAAWTS-----FAARRYGVSHPD 489
      : : : : :
Db 290 ALPAPQP-----LAQAGFHPDAPAPFAGIADYLAWGASRWASDAPRIAFILPRGAIADAQ 345
      : : : : :
QY 490 AGAAMRLLRSVYNSGACGHRSPVRRPSPLOMNTSIWYNSDVFBEAMR-----541
      : : : : :
Db 346 TGAIDELRRS-----ERHGOAPL-----AVWFDDSDP-EALRKSFAQADV 385
      : : : : :
QY 542 LLTSPASLATSAPARYDLDLTTRQAVQELVSLYEE-----ARSAYLSKELASL-- 591
      : : : : :
Db 386 QALVVLQHLQNGPAPARAFALD--VPLQTLGYRDGNEDWLAAAGVAPRTVAALFG 442
      : : : : :
QY 592 -----LRAGVLAYPELPAIDEVL-----ASDSRFLGSMV- 622
      : : : : :
Db 443 MPETWMSDPLVISALENGEPELWAG--QAEALDKLDRLRLRLPLVADKHALMFMNH 500
      : : : : :
QY 623 -----EQAPAAVSEAEADFEQNSRYQLTLMGPEGNILDYAN 660
      : : : : :
Db 501 PEGENVAASHLVNYPASLARGELRAAGYRVATSD-----ESALIDTAQ 545
      : : : : :
QY 661 KQLAGLVANYTTPRRRLFLBALV-DSVAGGIPFOQH 695
      : : : : :
Db 546 RLGG---YYRPQ---TLDALYRDGLASLPIDAY 574
      : : : : :
```

## RESULT 15

```
US-09-252-991A-33024
; Sequence 33024, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107136.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 33024
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-33024
```

Query Match 2.7%; Score 107.5; DB 4; Length 405;

Best Local Similarity 35.6%; Pred. No. 0.1; Mismatches 47; Conservative 9; Indels 19; Gaps 8;

```
QY 8 AAVGVLLAGA---GGA---GDEAREAAVRLVATLGGPAPADFSVSEBALAKP 60
      : : : : :
Db 251 AAGGAVLARGCWRKRGAAQRPGRRRPPARAGW--RAVGKPVAGVA---RRLAGEP 305
      : : : : :
QY 61 GLD--TYSLGGGGAARVVRGS-TGVAAAAGLHRYLRDFCGCHVAMSGSLRPLPAV 117
      : : : : :
Db 306 VRQAPAPALGGGAGARALVREGELAGATRHAGAR--GTHRRPRA--GLPQASDV 361
      : : : : :
QY 118 PGELEATPNRY 129
      : : : : :
Db 362 PGSLVFRPRRH 373
      : : : : :
```

Search completed: February 13, 2004, 16:18:02  
Job time : 26 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2004, 16:15:30 / Search time 56 Seconds

(without alignments)  
2778.054 Million cell updates/sec

Title: US-09-836-613-2

Sequence: 1 MEAAVAAGVAVLLAGAGC.....VDLAKKIFLKYPGWVAGSW 743

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 45 summaries

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pdp:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pdp:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pdp:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pdp:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pdp:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pdp:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pdp:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pdp:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pdp:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pdp:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pdp:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09D\_PUBCOMB.pdp:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pdp:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pdp:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pdp:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pdp:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pdp:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pdp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3939	100.0	743	11	US-09-836-613-2
2	2120	53.8	402	12	US-10-029-386-33001
3	832.5	21.1	728	15	US-10-156-761-9552
4	827.5	21.0	1038	15	US-10-156-761-13515
5	147.5	3.7	101	12	US-09-864-408A-3432
6	121	3.1	4342	9	US-09-815-242-5107
7	118.5	3.0	469	15	US-10-156-761-9695
8	118.5	3.0	871	15	US-10-367-019-7
9	118.5	3.0	1351	15	US-10-367-019-1
10	118.5	3.0	1610	15	US-10-367-019-4
11	118.5	3.0	1665	9	US-09-858-664A-2
12	118.5	3.0	2596	15	US-10-307-019-6
13	118.5	3.0	2630	14	US-10-077-130-2
14	118.5	3.0	7968	14	US-10-077-130-5
15	117.5	3.0	497	12	US-09-820-788-4

16	117.5	3.0	497	12	US-10-313-963A-60	Sequence 60, Appli
17	117.5	3.0	767	15	US-10-156-761-7798	Sequence 7798, Ap
18	115.5	2.9	3668	12	US-10-402-842-4	Sequence 4, Appli
19	114	2.9	473	12	US-10-368-493-10510	Sequence 10510, A
20	114	2.9	823	12	US-10-368-493-10184	Sequence 10184, A
21	113.5	2.9	567	12	US-10-368-493-16185	Sequence 16185, A
22	113.5	2.9	619	12	US-10-368-493-15803	Sequence 15803, A
23	113.5	2.9	771	12	US-10-368-493-15435	Sequence 15435, A
24	113	2.9	717	12	US-10-368-493-17769	Sequence 17769, Ap
25	113	2.9	7257	14	US-10-014-717-5	Sequence 5, Appli
26	109.5	2.8	1377	9	US-09-815-242-10384	Sequence 10384, A
27	109.5	2.8	1377	12	US-10-287-274-567	Sequence 467, App
28	108.5	2.8	796	15	US-10-156-761-14345	Sequence 14345, A
29	106.5	2.7	4498	15	US-10-156-761-5905	Sequence 19405, Ap
30	106	2.7	308	10	US-09-738-626-5728	Sequence 6728, Ap
31	106	2.7	310	9	US-09-826-909-2	Sequence 2, Appli
32	105.5	2.7	527	12	US-10-368-493-10548	Sequence 10548, A
33	105	2.7	4150	12	US-09-808-880-2	Sequence 2, Appli
34	104.5	2.7	726	11	US-10-368-493-23643	Sequence 23643, A
35	104	2.6	581	9	US-09-799-875-11	Sequence 11, Appli
36	104	2.6	581	9	US-09-757-982-8	Sequence 8, Appli
37	104	2.6	654	10	US-09-909-320-177	Sequence 177, App
38	104	2.6	654	10	US-09-909-0888-177	Sequence 177, App
39	104	2.6	654	10	US-09-905-281A-177	Sequence 177, App
40	104	2.6	654	10	US-09-902-853-177	Sequence 177, App
41	104	2.6	654	10	US-09-907-824-177	Sequence 177, App
42	104	2.6	654	10	US-09-907-841-177	Sequence 177, App
43	104	2.6	654	11	US-09-904-011-177	Sequence 177, App
44	104	2.6	654	11	US-09-906-742-177	Sequence 177, App
45	104	2.6	654	11	US-09-906-838-177	Sequence 177, App

## ALIGNMENTS

RESULT 1  
US-09-836-613-2  
Sequence 2, Application US/09836613  
Publication No. US20030039643A1  
GENERAL INFORMATION:  
APPLICANT: HOBWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;  
WEBER, BIRGIT; BLANCH, LINNIE; ANSON, DONALD STEWART  
TITLE OF INVENTION: SYNTHETIC MAMMALIAN  
N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES  
ENCODING SAME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: NIXON PEARBODY LLP  
STREET: 990 STEWART AVENUE  
CITY: GARDEN CITY  
STATE: NEW YORK  
COUNTRY: UNITED STATES  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/836,613  
FILING DATE: 17-Apr-2001  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00747  
FILING DATE: 22-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: POKALSKY, ANN R.  
REGISTRATION NUMBER: 34,697  
REFERENCE/DOCKET NUMBER: 2249/104  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516 742 4343  
TELEFAX: 516 742 4366  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 743 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 261  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 272  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 435  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 503  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 513  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 526  
FEATURE:  
NAME/KEY: Potentially-glycosylated Asn site,  
LOCATION: 532  
SEQUENCE DESCRIPTION: SEQ ID NO: 2  
US-09-836-613-2

Query Match 100.0%; Score 3939; DB 11; Length 743;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEAVAAVAAVGVLLGAGAGAGDEARBAARALVARLIGPPADPFVSVERALAAAF 60  
DB 1 MEAVAAVAAVGVLLGAGAGAGDEARBAARALVARLIGPPADPFVSVERALAAAF 60  
QY 61 GUDTYSIGGGGAARVRVRSSTGVAAAAGLHRYLRDFCGCHVAVSSGQLPPLPAVGE 120  
DB 61 GUDTYSIGGGGAARVRVRSSTGVAAAAGLHRYLRDFCGCHVAVSSGQLPPLPAVGE 120  
QY 121 LTRATPRRYRYONVCCOSYFVWMDARWEREDIMMANGINLALAMSGEAIWQRYL 180  
DB 121 LTRATPRRYRYONVCCOSYFVWMDARWEREDIMMANGINLALAMSGEAIWQRYL 180  
QY 181 ALGLTOAINEFTGPAFLAMGMGNLHTWGPJPSWHIKQLYLQHRVLDQMSFGMT 240  
DB 181 ALGLTOAINEFTGPAFLAMGMGNLHTWGPJPSWHIKQLYLQHRVLDQMSFGMT 240  
QY 241 VLPAPAGHVEATRVPPQVNTKMGSGHFNCSYCSFLAPEDIFPLIGSLFREL 300  
DB 241 VLPAPAGHVEATRVPPQVNTKMGSGHFNCSYCSFLAPEDIFPLIGSLFREL 300  
QY 301 KEETGTHIYADTFENEMQPPSSSEPSYLAATTAVSAMTAVIDEAVMLQGMFOHQPF 360  
DB 301 KEETGTHIYADTFENEMQPPSSSEPSYLAATTAVSAMTAVIDEAVMLQGMFOHQPF 360  
QY 361 WGAPOIAVAVGAVPRGGLVLDLFAESQPYRTTASFOGQPTWMLAHNGHGLFGL 420  
DB 361 WGAPOIAVAVGAVPRGGLVLDLFAESQPYRTTASFOGQPTWMLAHNGHGLFGL 420  
QY 421 EAVVAGGEAARLFENSTMGVGMABEGISONEVYVSLMAELGMRKDPVDLAAWTSFAA 480  
DB 421 EAVVAGGEAARLFENSTMGVGMABEGISONEVYVSLMAELGMRKDPVDLAAWTSFAA 480  
QY 481 RRYGVSHPDGAARLLRSVYVNCSGACGHRSPVLRPSIQNTSTWYNSDVFEAM 540  
DB 481 RRYGVSHPDGAARLLRSVYVNCSGACGHRSPVLRPSIQNTSTWYNSDVFEAM 540  
QY 541 RLILTSAPSLATSPAFRYDLDLTROAVQELVSLYEARSAYLSKELSLFAGVLA 600  
DB 541 RLILTSAPSLATSPAFRYDLDLTROAVQELVSLYEARSAYLSKELSLFAGVLA 600  
QY 601 ELIPALDEVLASDSRFLGSMLEQARAASVSEADTFEONSRVQLTMGPBGNILDIYAN 660

DB 601 ELIPALDEVLASDSRFLGSMLEQARAASVSEADTFEONSRVQLTMGPBGNILDIYAN 660  
QY 661 KQLAGLVANVTTPMRFLFLBALVDSVAQGIPOOHQDFKRVFQLECAFVLSKORVPSQPR 720  
DB 661 KQLAGLVANVTTPMRFLFLBALVDSVAQGIPOOHQDFKRVFQLECAFVLSKORVPSQPR 720  
QY 721 GDTVDLAKKIFLKYPGVAGSW 743  
DB 721 GDTVDLAKKIFLKYPGVAGSW 743

RESULT 2  
US-10-029-386-33001  
Sequence 33001, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Hanzel, David R.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: AEMTCA-X-2  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 3428  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 33001  
LENGTH: 402  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO U34879.1  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.74  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.93  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6  
OTHER INFORMATION: SWISSPROT HIT: P54802, EVALUATION 0.00e+00  
US-10-029-386-33001

Query Match 53.8%; Score 2120; DB 12; Length 402;  
Best Local Similarity 100.0%; Pred. No. 6; 9e-185;  
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 DTEAVMLIQGMFOHQPFQWGPQAIKAVLGAAPRGGLVLDLFAESQPYRTTASFOGQ 401  
DB 1 DTEAVMLIQGMFOHQPFQWGPQAIKAVLGAAPRGGLVLDLFAESQPYRTTASFOGQ 401  
QY 402 FIVCMLEHFGGNHGLFGALBAVNGGPEAARLFPSITMGVGMABEGISONEVYVSLMAEL 461  
DB 61 FIVCMLEHFGGNHGLFGALBAVNGGPEAARLFPSITMGVGMABEGISONEVYVSLMAEL 461  
QY 462 GMRKDPVDLAAWTSFAARVGVSHPDGAARLLRSVYVNCSGACGHRNSPLVRP 521  
DB 121 GMRKDPVDLAAWTSFAARVGVSHPDGAARLLRSVYVNCSGACGHRNSPLVRP 521  
QY 522 SLOMNTSTWYNSDVFEAMRLLTSAPSLATSPAFRYDLDLTROAVQELVSLYEARS 581  
DB 181 SLOMNTSTWYNSDVFEAMRLLTSAPSLATSPAFRYDLDLTROAVQELVSLYEARS 581  
QY 181 SLOMNTSTWYNSDVFEAMRLLTSAPSLATSPAFRYDLDLTROAVQELVSLYEARS 581  
QY 582 AYLSKELSLFAGVLAAYELLPALDEVLASDSRFLGSMLEQARAASVSEADTFEON 641  
DB 241 AYLSKELSLFAGVLAAYELLPALDEVLASDSRFLGSMLEQARAASVSEADTFEON 641  
QY 642 SRVQLTMGPBGNILDIYANVAVVYVPRRLFLALVDSVAQGIPOOHQDFKRV 701  
DB 301 SRVQLTMGPBGNILDIYANVAVVYVPRRLFLALVDSVAQGIPOOHQDFKRV 701  
QY 702 FQLECAFVLSKORVPSQPRGDTVDLAKKIFLKYPGVAGSW 743  
DB 361 FQLECAFVLSKORVPSQPRGDTVDLAKKIFLKYPGVAGSW 743

RESULT 3  
 US-10-156-761-9552  
 ; Sequence 9552, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156, 761  
 ; PRIOR FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 9552  
 ; LENGTH: 728  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-9552

Query Match 21.1%, Score 832.5; DB 15; Length 728;  
 Best Local Similarity 31.9%; Pred. No. 8.3e-67;  
 Matches 239; Conservative 106; Mismatches 338; Indels 67; Gaps 21;

12 VLLAGGAGGADGAREAAVRLVALLGPGPADSVSVERALAAK-----PG 61  
 7 VLLTLLAAGSAG-----AAVACDPASADG-APAGSAAATARRLLPRHRQLTFRAGG 59  
 62 LDYSLGGGGAARVVRGSGTGAAGLHRYLDFCCGHVAMSGSGLRPRPAPVAGEL 121  
 60 RDTFOV-SGRTGRVTVVGGTATQTLGLNMYLRINADINMAGROLRLPRALPGLAGTV 118  
 122 TEAPNRYRYQNVCTGYSFVWMDMAREREIDMMALNGINILAMSGOEALMORYLA 181  
 119 TRRANVRERFALNTNDGTYGPHDWTYMERELDLVLAHGCNEVLVGTGADLHHRVFOE 178  
 182 LGLTQAEINEFTGPAFLAMGRMGNLHTWDGRLPPSHIKOLYLQHRVLDQMSFGMT 241  
 179 FGTYDELRKMIKPPRAPHQMWLLQNLGAFRVPVSGQLDARALNRIARLRLHGMT 238  
 242 LPAPAGHVEAVTVFPOVNVTKMGSGHFNCSYSCSFLAPEDPIPTIGSLFLRELK 301  
 239 FPGYFGTVPPGAFDRNAGAHVPOGTW---MGFARPDWLDPRTEHFRVAAAFYRQDE 294  
 302 EF-GTDDHYGADTENEMOPSSSEPSYLAATTAYEAMTAVDTEAVVLLQGMFOHOPF 360  
 295 MFGASSTRYKMDLHGGSPGDV--VODAKGYERLRAHAPAVVILGM--QHNP- 349  
 361 WGPAPQIRAVLGAVPRGRLLVLDLFAESQPVYT-RTASFGQOPFTICMLHNGNGLFGA 419  
 350 -----RAIVDAVYKDRMLVVDGLCDRFPKVTDEADMGHGPYAFSGIMVFGG-HTLGA 402  
 420 LEAVNGGEAARLF-----NSTMTVGTMAREGISQNEVVYSMAELGMRKDPVPLAA 473  
 403 -----NTPDMASLTERMTFRPGSTLRGVALPEADNPPAFALFSELARREGDL-DLRA 466  
 474 WTSFPAARRYGVSHPDGAAMRLILRSYNGSGEACRGNRSPLY-RRPSLQMTSTIWN 532  
 457 WFAWARSRGYGRDPHAEADWILRRTAYGTTRADSMSEGADLFGARPSLAATKASWS 516  
 533 -----RSDVE-AMRLILTSABSLATSPARRYLLDLTROAVELVSLTYEERKASYLS 565  
 517 PKRLRYRREEREPALGELLKVRPGLRGSARRDLLVARQALSNRRSVLLPQIRTYEA 576  
 586 KELASLIRAGGV-LAYELLPALDEVLASDSRFLGSMLEQARAAYSEADFYEQSRX 644

DB 577 KOTAREDLTGWMLA--LMDLLEALLATDSRLLGRWVADARAWGASAEERDLAYDAL 634  
 QY 645 QLTW-----GPEGNILDYANKOLAGVANYYPFRRLFLFALVDSVAGIPFQOHOPDN 700  
 DB 635 LITWGTGAGAGLADYANRREWAGLVGGLYLRKSTYFAELRSASREERTKTKDW---- 691  
 QY 701 VFQLEQAFVLSKORYPSQPRGDTVLAKKI 730  
 DB 692 -PALEDRTWRNPGGLATRPDTGYQAAVAV 720

RESULT 4  
 US-10-156-761-13515  
 ; Sequence 13515, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156, 761  
 ; PRIOR FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 13515  
 ; LENGTH: 1038  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-13515

Query Match 21.0%, Score 827.5; DB 15; Length 1038;  
 Best Local Similarity 29.9%; Pred. No. 4.1e-66;  
 Matches 222; Conservative 121; Mismatches 359; Indels 41; Gaps 15;

QY 3 AVAANAAGVLLLAGAGGAGD-EAREAAVAVRLVALLGPGPADSVSVERALAAK 61  
 DB 15 AIGLGTGTTCTALPAHATSPGEGPLDTPABASALNRL-PCHAQFRSLT---VGRTRG 70  
 QY 62 LDYSLGGGGAARVVRGSGTGAAGLHRYLDFCCGHVAMSGSGLRPRPAPVAGEL 121  
 DB 71 RDRFRV-TGGTRRIVSGTTPAVLLITGVHWYLVKVCAGHLANNGQLDLPRLPAPAPL 129  
 QY 122 TEAPNRYRYQNVCTGYSFVWMDMAREREIDMMALNGINILAMSGOEALMORYLA 181  
 DB 130 ERSTALSHRFALNDNDGTAPAYADMSYWEHQIDLLAHGCNEVAVVINGTEAVRYLKD 189  
 QY 182 LGLTQAEINEFTGPAFLAMGRMGNLHTWDGRLPPSHIKOLYLQHRVLDQMSFGMT 241  
 DB 190 FGYSDTEARAWLPASHOPFWMLQNLISGYSGLSPELIABRAGLRRICDRIRALGMAV 249  
 QY 242 LPAPAGHVEA-VTVFPOVNVTKMGSGHFNCSYSCSFLAPEDPIPTIGSLFLREL 300  
 DB 250 LPEYIGHVPKGVENRGDAHVPOGTWGF-----ERDWDLPRTASPAAYAKSYRRQK 305  
 QY 301 KEFGTDHYGADTENEMOPSSSEPSYLAATTAYEAMTAVDTEAVVLLQGMFOHOPF 360  
 DB 306 DVFQGAARHKNMOLLHGGTAGDVP--VGAARGVEKALQAAHPGATVWILGAEANPLP- 361  
 QY 361 WGPAPQIRAVLGAVPRGRLLVLDLFAESQPVYT-RTASFGQOPFTICMLHNGNGLFGA 419  
 DB 362 -----ALDLAIDKKKMLIVDGVSDRYTSYDRKEDMGGTPYAFGTTIPNFGGRTTIGAR 414  
 QY 420 LEAVNGGEAARLFNPSNTMTVGTMAREGISQNEVVYSMAELGMRKDPVPLAAVTSFA 479  
 DB 415 AHUMNEKFAWMDKAGSALAGTAYLPEAADRDPAFALFSELAWAGAKI-DRAAWFSSYA 473

```

CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 1410
SOFTWARE: PasteSeq for Windows Version 4.0
SEQ ID NO 5107
LENGTH: 4342
TYPE: PRT

```

	Query March	3.1%;	Score 121;	DB 9;	Length 4342;
	Best Local Similarity	21.4%;	Pred. No. 1.4;		
	Matches 178;	Conservative 101;	Mismatches 318;	Indels 236;	Gaps 44;
Cy	4 VAVAA-----AVGVLLIAGAGA-----AGDEAPPAAYRA-TVARLL	40			
Db	1181 VALCAERSPOLLVGLLIVRAGGAYVLPDDEYSERLAYMLAAGVELLLTQAHIFERL-	1239			
Cy	41 GPG-----PAADPSVSVERLLAAKPSGIDTSLSGGGAARVRSGSTVAAAAGHRRRLRD	95			
Db	1240 -FGABGVTFICLDSLKLDNMPGAPGLHLH--GDNLAYVYTTSGSTQPKVG-----	1289			
Cy	96 FCGCHVAMSGSGLRLPRPLPAVPG--LTETATPNRYRYQYQVCTQSYSFVMDMARWERE	153			
Db	1290 --NTHALLA-ERLQWQATYTLDDGVLMQKAP-----VSFDVSW--E	1348			
Cy	154 IDWMAALNGINLALMSQ-----EAIQRYVLAQLTQAEINEPFTGAFILAMQMG	205			
Db	1329 CFMWLVGGCRVTLAAPEHNDPAVLVELVSGVTTLHFVPELQHTIDEGVAACSLR	1388			

Db	1389	RLTEGGGALPRF-----	LNRRVLQR-----	LPAAALNHRVQPTER--	AINVT--	1426
Qy	266	GSWGHENCSSYCSFLAPED---	PIFPITGSLFRELKEFGTDH	YGADTENEM---		317
Db	1429	---HMOC-----	RAEDGERSPIRPLGNVCVYLDAEF--	NLLPAGVAGELCIGG	1473	
Qy	318	-----QPPSEPSYLAATNAV	VEAMTAVDTEAVMLQGL-----	FQHPQPMG-	362	
Db	1474	LGILRGYIGRPAUSAERFVAD	PFSAGGERLRTIGDRARMADGVL	BYLGRIDQVKRGF	1533	
Qy	363	---PAQIRAVGVAVP---RGR	LVLDFEAFESQPY--YTRTAS	FOGQEPFWCMHNFGNH	414	
Db	1534	RIBEEBICARLLAPPGVAGAVV	IRBQVAGSQLVGYTTGAVGA	BAEN-----EQNQ	1584	
Qy	415	GLFALBAVNGGPPAARLPFN	STNVGTGMAREGISONEVYTS	LMAELGMK---DPYD	470	
Db	1585	RLRALAQ-----ELPEY	MVPTQIMRLACMPLGPS--	GKLTJRALPEYVQOREHVEPRTE	1638	
Qy	471	L-----AAWTSFARRYG-----		YSHDGAAMRLLRSYVNS--	505	
Db	1639	LORRIATWSVTLGLPRVGLR	DDFFELGSHSLATRVSTRQAC	DVELPRALFEASEL	1698	
Qy	506	-----GEACRGHNSPLVR	PPSLQMTSTIWNSDYVEAWR	LLTSPASLATSBAF	556	
Db	1699	EAFCEQVRAQAARTDSHAI	RIRIDREQRPPLYSQRMWFLQ	-----EPD---SPAY	1751	
Qy	557	RY-----DLIDITR--QAVQ	VELSYTYEBARSATYLSKEL	ASLALRA---GGV-LAYELL	603	

Db 1752 NVGGLARLSPGLDVARFEALQALVQ-RHETLRTTPSPVGVQVRVHGGGLHMDWQDF 1810  
 Qy 604 PALD-----EVLASRSRFLGSMLEQA---RAAVSEKADPFYQNSRYQULTLMGPE 652  
 Db 1811 SALDRDSRQOHLQTLA-DSEAHKPFDESGPLLEVCVKAERHEHYLVTLHHTV---TE 1866  
 Qy 653 GNILDYANKQIAGVANYTYTPRWLFLEALVDSVAQGIPEQOQOFQKXVQLE 705  
 Db 1867 GWANDIFARELGALYEAFLDLR-----ESPLLEPLPVQ-YLDYSVMQRE 1908

RESULT 7  
 US-10-156-761-9695  
 ; Sequence 9695, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156, 761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 9695  
 ; LENGTH: 469  
 ; TYPE: PR1  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-9695

Query Match 3.0%; Score 118.5; DB 15; Length 469;  
 Best Local Similarity 22.1%; Pred. No. 0.078;  
 Matches 126; Conservative 49; Mismatches 167; Indels 229; Gaps 32;  
 Qy 208 HTWDQPLPSPWHIKQLYLOHRLVDM-----RSFMTPLVPAFA 246  
 Db 14 HEMKALPP-----LIRPTVDELARAAALVARRRGDARLRVGTSGGTPLMLISV 66  
 Qy 247 GH-VBEAVTRVFPVQVNTKMGWGHFNCSYSCSFLAEPEDPIPIIGSLFLRELKEF-G 304  
 Db 67 GHGTRHALVAGPHANE-----PVGATALRLAERYLAG 100  
 Qy 305 TDHITYGATFEMQPPSEPSYLAATTAVYEAMTAVDTEAVMLLQGL-----FQ 355  
 Db 101 PGPNEGAD-----ATNNLLICDDPGARNEGMLSGPYTLGHYFR 140  
 Qy 356 H-----QPFMGPAQIRAVLGAVERGLLVLDLFAESQPVYTRTASFOGQPIWCM 407  
 Db 141 HFFRRGFLQEFWMLPDGAERA---ALPSTRAL-LDLQDEL-----REFFOCSL 184  
 Qy 408 H--NNGC-----NHLFGALNAVNGGPRAR-----LFPNSIMVGTGMAPEGISQNEV 454  
 Db 185 HGVVDGGGFFVELTODLPGLAQVRAG--TAAKFGIPREIGPYDTLWPPPLG-----AV 235  
 Qy 455 YSIMELEMRKDPV--DLAAWTS-----FAARRGV-----S 486  
 Db 226 YRI-----PPRRGDLLAAITFAVESTWPHQRYGTVAVVEAPMVGAAVEDGS 286  
 Qy 487 HP-DGAAMRLILRBVNVCSG--EACRGHNSPLVRPSLOMTSITWNRSDVEARLL 543  
 Db 287 PPGDADAVLRAVSRLRDTGLIEDILARIRPRLAATPD---TASLAPVDD---Y 336  
 Qy 544 LTSAPSLATSPAFRDLDLTFQAVQEL-----VSLYEERASVYSKELASLLRA 594  
 Db 337 LLVGVGLADS--WDDTFTDGTGRPLPLNTARLTLGSGRRVALRTAGLLRLAT---A 391

Qy 595 GGVLAYELLPALDEVILASDSRFLGSMLEQAARAAVSEADPFYQNSRYQULTLMGPEGN 654  
 Db 392 AGAASDVLPFLDR-----LIDWCADYR-----DGYAR-----WVPVGH 427  
 Qy 655 ILDYANK-----QIAGVANYTYT---PRW 675  
 Db 428 QVEYQTRVVLAAFAELAGRYAPVRSHSGEPRW 458

RESULT 8  
 US-10-307-019-7  
 ; Sequence 7, Application US/10307019  
 ; Publication No. US20030108533A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zeng, Wenlin  
 ; APPLICANT: Station, Lawrence  
 ; APPLICANT: SCIOS, INC.  
 ; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION  
 ; FILE REFERENCE: SCIOS.021DVL  
 ; CURRENT APPLICATION NUMBER: US/10/307,019  
 ; CURRENT FILING DATE: 2002-11-26  
 ; PRIOR APPLICATION NUMBER: 09/548,473  
 ; PRIOR FILING DATE: 2000-04-13  
 ; PRIOR APPLICATION NUMBER: 60/129,552  
 ; PRIOR FILING DATE: 1999-04-16  
 ; NUMBER OF SEQ ID NOS: 9  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 7  
 ; LENGTH: 871  
 ; TYPE: PR1  
 ; ORGANISM: Homo sapiens  
 US-10-307-019-7

Query Match 3.0%; Score 118.5; DB 15; Length 871;  
 Best Local Similarity 18.6%; Pred. No. 0.2;  
 Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;  
 Qy 19 GCAQDEARBAANALVALRL-LGQGPAD---FSVSRALAAAPGLD-----TY 65  
 Db 21 GSAQSSQSSQSVSLRGSSQVGTSPSLDAGCMQEAADLSDSTPLQRPQEQATMKXF 80  
 Qy 66 SLGG-GGAARVVRGSGTVAAAGLHRYLRDCG-----CHVAGSGQLR-----109  
 Db 81 SLGGRGVAGVAGYGFARFGDAG-----GMLGGPMWARIAMVSGSEEEQEARAE 134  
 Qy 110 -----LP---RPLAVPGELTATPRRYQNVCTQSYSPWMMARWE 151  
 Db 135 SQSEEQEARASPLPQVSARFVPEVGRAPTRSSP-----EPTWE 175  
 Qy 152 -----REL-DWMLNGINLAL-----AMSGQEAIMQRYIAL-----182  
 Db 176 DIGQVSLVQIRLDSGDAEAADTISLDSVDAVYLNLSLYIKLIPFPMFRVYKSA 235  
 Qy 183 -----GLTQAEINEF-----FTGP-----ATLWGRMGNLHTWGP 213  
 Db 236 QEPPEPMAEEELAEPRPTWPGELGPHAGLETIESEEDVDALAEAAVGRKRWSP 295  
 Qy 214 LPPSWHILKQLYLOHRLDQMRSGMTPVLPAPAGHV-----PEAVTVFPQVNTKMG 266  
 Db 296 SRSLLFPFG--RHLPDPAELGLRERYKASVEHISRLIKRPSGLKEGSP---RKRP 349  
 Qy 267 SWGHFNCSYSCSFLAEPDPIPIIGSLFLRELKEFGTDHLYGADTFEMQPPSEPSY 326  
 Db 350 GLASFRLSGKSWDRAP-----FLREL-----SDETVV 378  
 Qy 327 LAAATTAVENTAVDTEAVMLLQGLPFQHQQFMPGPAIRAVLGVPRG--RLVLDLF 384  
 Db 379 LGGSVTLAQVSAQPAQATWCKDGPDE-----SSKRLISATLTKNFDLLITLVV 430  
 Qy 385 AESQPVYTRTASFOGPIFCMLHNFQGNHGLFGALNAVNGGPEARLFPNSIMVGTGMA 444  
 Db 431 AEDLGYIT-----GVSNALGVTTTGVLR-----KMR--PSSS-----463

```

QY 445 PEGISQNEVVYSLMAELGMRKDPVPD-----LAAM--VTSFAARRGVSHPDAGAA 493
DB 464 -----PCPDIGEYADGVLLVWKPVESYGPVTYIVQCSLEGGS 501
QY 494 WRLLRSVYNC-----SGEAC 509
DB 502 WTLASDIFDCCYLTSKLSRGTYTFRACVSKAGMPYSSPSEQVLLGGPSHLASEES 561
QY 510 RGNHRSPLYRPSLQMTSTIWNR--SDVFEAMRLLLTSAPSLATSAPFYDLIDLTR--Q 566
DB 562 QGRSAQPLPSTKTFAFOTQIORGRFSVRQCEKASGRALAAKTIYPHPKDKTAVLREYE 621
QY 567 AVOELVSLYYEARSAYLS--KELASLRAGVLAVELLPALDEVLASDSRFLIGSWLEQA 625
DB 622 ALKGLRHPHLAQHLAAVLSPRHLVLIJEL--CSGPELLPCLAE--RASYESSEVQDYLMQM 678
QY 626 RAAA-----VSEADPFEQN---SRQYLTIMGPEGNILIDYANKQLAGLVANYITPRML 677
DB 679 LSAQYILNQHILHLDLRSENMTITEYNLL-----KVVDLGNAGSLSQEKVLPSPDKFO 732
QY 678 FLEALVDSVACG---IPFOQHOPDKNVFOL--EQAFVLSKQRYSPQPRG 721
DB 733 YLEMAPLELGGQAVP-----QTDIMAGVTAFTMLSAEYVSSSE 774

RESULT 9
US-10-307-019-1
; Sequence 1, Application US/10307019
; Publication No. US20030108533A1
; GENERAL INFORMATION:
; APPLICANT: Zeng, Wenlin
; APPLICANT: Stanton, Lawrence
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
; FILE REFERENCE: SCIOS.021DVL
; CURRENT APPLICATION NUMBER: US/10/307, 019
; PRIOR FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/548,473
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/129,552
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1351
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-307-019-1

Query Match 3.0%; Score 118.5; DB 15; Length 1351;
Best Local Similarity 18.6%; Pred. No. 0.39;
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

QY 19 GGAAGDEAREAAVAVLVARL-LGGPAD--FSVSVERALAKPGLD-----TY 65
DB 501 GSAAGASSQVSSIRVSSQVGTGEPGSLDAEGWTOEAEDLSSTPTLQRPQOATWKRF 560
QY 66 SLGG-GGAARVRVGSVCVAAAAGLHRYLDFCG-----CHVAMSGSQLR----- 109
DB 561 SLGGRGYAGVAGYGTAFGGDAG-----GMLQGGWMAKRLMAVVSQSEEEQEARAE 614
QY 110 -----LP---RPLPAVGELEATPNNRYRYQVNCQSYSFVMDWARWE 151
DB 615 SQSEEQEARAESPLPOVSARPVVEGRAPTRSSP-----EPTPWE 655
QY 152 -----RET---DMMLINGINLAL-----AMSQGEATWQRYLAL----- 182
DB 656 DIGQVSLVQIRDSGDADADTISLDISEVDPAVLNLSLDYIKYLPFEPMIRKQPKSA 715
QY 183 -----GLTQAEINEF-----FTGP-----AFLAMGRMGMLHTWDG 213
DB 716 QPFPSPMAEBELAEFPFPPTWPGELGPHANGLETSESDVALLAEAVAGKRWSSP 775

```

```

QY 214 LPSPMHIKQLYLQHRVIDQMSFGMTPEVLPAFAGV-----PEAVTRVFPQVNVTKMG 266
DB 776 SRSLEHFRG---RLPLDEPAREGLRBRVYKASVHIRILKGRREGKEGCP--RKP 829
QY 267 SMGHNCSYSCSFLLAPEDPIFPITGSLFRELKEGTGHIIYADTFNMQPPSSPSY 326
DB 830 GLASFRLSGLXSWRAP-----TFPLREL-----SDETVV 858
QY 327 LAAATTAVEAMTAVDEAVWLQGMLEFOHPQCFWGAQIRAVAGVPRG--RLVLIDL 384
DB 859 LGQSTYLAQCVSAPPAQATMSKQAPLE-----SSSRVLISATLKNKQLLTLVV 910
QY 385 AESQPVYTRTASFQGPPIWCMNLNFGNHLFGALVAVNGGPPAARLFPNSTWVGMA 444
DB 911 AEDIGVYI-----CSVSNALGTVTTGVLR-----KAER--PSSS----- 943
QY 445 PEGISQNEVVYSLMAELGMRKDPVPD-----LAAM--VTSFAARRGVSHPDAGAA 493
DB 944 -----PCPDIGEYADGVLLVWKPVESYGPVTYIVQCSLEGGS 981
QY 494 WRLLRSVYNC-----SGEAC 509
DB 982 WTLASDIFDCCYLTSKLSRGTYTFRACVSKAGMPYSSPSEQVLLGGPSHLASEES 1041
QY 510 RGNHRSPLYRPSLQMTSTIWNR--SDVFEAMRLLLTSAPSLATSAPFYDLIDLTR--Q 566
DB 1042 QGRSAQPLPSTKTFAFOTQIORGRFSVRQCEKASGRALAAKTIYPHPKDKTAVLREYE 1101
QY 567 AVOELVSLYYEARSAYLS--KELASLRAGVLAVELLPALDEVLASDSRFLIGSWLEQA 625
DB 1102 ALKGLRHPHLAQHLAAVLSPRHLVLIJEL--CSGPELLPCLAE--RASYESSEVQDYLMQM 1158
QY 626 RAAA-----VSEADPFEQN---SRQYLTIMGPEGNILIDYANKQLAGLVANYITPRML 677
DB 1159 LSAQYILNQHILHLDLRSENMTITEYNLL-----KVVDLGNAGSLSQEKVLPSPDKFO 1212
QY 678 FLEALVDSVACG---IPFOQHOPDKNVFOL--EQAFVLSKQRYSPQPRG 721
DB 1213 YLEMAPLELGGQAVP-----QTDIMAGVTAFTMLSAEYVSSSE 1254

RESULT 10
US-10-307-019-4
; Sequence 4, Application US/10307019
; Publication No. US20030108533A1
; GENERAL INFORMATION:
; APPLICANT: Zeng, Wenlin
; APPLICANT: Stanton, Lawrence
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION
; FILE REFERENCE: SCIOS.021DVL
; CURRENT APPLICATION NUMBER: US/10/307, 019
; PRIOR FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/548,473
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/129,552
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1610
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-307-019-4

Query Match 3.0%; Score 118.5; DB 15; Length 1610;
Best Local Similarity 18.6%; Pred. No. 0.51;
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

QY 19 GGAAGDEAREAAVAVLVARL-LGGPAD--FSVSVERALAKPGLD-----TY 65
DB 760 GSAAGASSQVSSIRVSSQVGTGEPGSLDAEGWTOEAEDLSSTPTLQRPQOATWKRF 819

```

```

QY 66 SLGG-GGAARVRVSGTGVAAAGLHRYLADFCG-----CHVAMSGSOLR-----109
DB 820 SLGGRGYAGVAGYGTAFPGDAG-----GMLGGGMARIAMAVSQQSEEEGEARAE 873
QY 110 -----LP-----RPLPAVPGELTEATPNRYRYQNVCTQSYSFVMMDMARME 151
DB 874 SQSEGEQEARABESLPQVSARVPVEGRAPTRSSP-----EPTPWE 914
QY 152 -----REI--DMALNGINLAL-----AMSGEAIHQRYLAL-----182
DB 915 DIGQVSLVQIRDLSDGDAEADTISLDISEVDPAVNLSDLDYDIKYLPEEFMIFRKVPKSA 974
QY 183 -----GLTQAEINEF-----FTGP-----AFLAMGRMGLHTWDGP 213
DB 975 QPEPPSPMAEELAEFFPEPTWMPGELGPHAGLEITESEVDALLAEAVGRKKMSSP 1034
QY 214 LPPSMHILKQLYQHRVLDQMSFGMTVLPAPAGHV-----PEAVTRVPQVNVTKMG 266
DB 1035 SRSLEHFPFG--RHLPDEPAELGLRERVKASVEHISILKGRPEGLEKGGPP--RKKP 1088
QY 267 SMGFHNCYSQSGFLAPEDPIFPIIGSLFLRELIKERGTDIHYGADTFNEMQPPSEPSY 326
DB 1089 GLASFRLSGLKSWDAP-----FTLREL-----SDETIV 1117
QY 327 LAAATTAVYEAMTAVTEAVWLLOGMLFOHQPOFGMPAQIRAVLGAVERG--RLVLVDLF 384
DB 1118 LGQSVTLACQVSAQPAQATWSKDGAPE-----SSSRVLISATLKNFOLLITLVV 1169
QY 385 AESQPVYTRTASFGQGPFIWCMLNHPGNGHGLFGALEAVNGGPEARLFPNSTWVTGMA 444
DB 1170 AEDLCVIT-----CQSNALGTVTITGVLR-----KAER--PSSS-----1202
QY 445 PEGISONEVVYSIMAEIGMRKDPVD-----LAAM--VTSFPAARYGVSHPDAGAA 493
DB 1203 -----PCPDIGEVYADVGLVWKPVEISYGVITYIVQCSLEGGS 1240
QY 494 WRLLRSVYNC-----SSEAC 509
DB 1241 WTTLASDIFDCCYLTSKISRGTYTFTACVSKAMGPPSSSESVOLLGAPSHLASEES 1300
QY 510 RGHNSPLVRRPSLOMNTSIWYNR--SDVEAMRLLLTSAFSLATSPAFRYDLDTLR--Q 566
DB 1301 QGRSAQPLPSTKTPAFQIQGRFVSVRQCEKASGRLAAKIIPIHPKDXTAVLREYE 1360
QY 567 AVQELVSLYEBAKSAIYS--KELASLLRAGGTLAYELLPALDEVLASDSRFLGSMLEQA 625
DB 1361 ALKGLRHPHLAQLHAAYLSPRHVLITLLEL--CSGPELLPCIAE--RASYSSESVQDYLMQM 1417
QY 626 RAAA-----VSEAEADFEYQN--SRYQULTMGPEGNILDYANKQLAGLVANYTTPRWRL 677
DB 1418 LSAQVYLHNQHLHLDLASENNIITEYNLL-----KVVDLGNAGSLSQEKVLPBDKFXD 1471
QY 678 FLEALVDSVAQG--IPFQGHQFDKXVQL--EQAFVLSKORPSQPRG 721
DB 1472 YLETMAPELLBQGAVP-----QTDIMAIGVTAFIMLSAEYVSSSEG 1513

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-858-664A-2
Query Match      3.0%; Score 118.5; DB 9; Length 1665;
Best Local Similarity 18.6%; Pred. No. 0.54;
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

QY 19 GGAADDEAREAAVALVARL-LGEPDAD--FSVSERALAAKPGLD-----TY 65
DB 815 GSASQASSSQVSLRVGSSQVTEBPGLDAGMTQAEEDLSSTPTLQRPQEQVTMRKF 874
QY 66 SLGG-GGAARVRVSGTGVAAAGLHRYLADFCG-----CHVAMSGSOLR-----109
DB 875 SLGGRGYAGVAGYGTAFPGDAG-----GMLGGGMARIAMAVSQQSEEEGEARAE 928
QY 110 -----LP-----RPLPAVPGELTEATPNRYRYQNVCTQSYSFVMMDMARME 151
DB 929 SQSEGEQEARABESLPQVSARVPVEGRAPTRSSP-----EPTPWE 969
QY 152 -----REI--DMALNGINLAL-----AMSGEAIHQRYLAL-----182
DB 970 DIGQVSLVQIRDLSDGDAEADTISLDISEVDPAVNLSDLDYDIKYLPEEFMIFRKVPKSA 1029
QY 183 -----GLTQAEINEF-----FTGP-----AFLAMGRMGLHTWDGP 213
DB 1030 QPEPPSPMAEELAEFFPEPTWMPGELGPHAGLEITESEVDALLAEAVGRKKMSSP 1089
QY 214 LPPSMHILKQLYQHRVLDQMSFGMTVLPAPAGHV-----PEAVTRVPQVNVTKMG 266
DB 1090 SRSLEHFPFG--RHLPDEPAELGLRERVKASVEHISILKGRPEGLEKGGPP--RKKP 1143
QY 267 SMGFHNCYSQSGFLAPEDPIFPIIGSLFLRELIKERGTDIHYGADTFNEMQPPSEPSY 326
DB 1144 GLASFRLSGLKSWDAP-----FTLREL-----SDETIV 1172
QY 327 LAAATTAVYEAMTAVTEAVWLLOGMLFOHQPOFGMPAQIRAVLGAVERG--RLVLVDLF 384
DB 1173 LGQSVTLACQVSAQPAQATWSKDGAPE-----SSSRVLISATLKNFOLLITLVV 1224
QY 385 AESQPVYTRTASFGQGPFIWCMLNHPGNGHGLFGALEAVNGGPEARLFPNSTWVTGMA 444
DB 1225 AEDLCVIT-----CQSNALGTVTITGVLR-----KAER--PSSS-----1257
QY 445 PEGISONEVVYSIMAEIGMRKDPVD-----LAAM--VTSFPAARYGVSHPDAGAA 493
DB 1258 -----PCPDIGEVYADVGLVWKPVEISYGVITYIVQCSLEGGS 1295
QY 494 WRLLRSVYNC-----SSEAC 509
DB 1296 WTTLASDIFDCCYLTSKISRGTYTFTACVSKAMGPPSSSESVOLLGAPSHLASEES 1355
QY 510 RGHNSPLVRRPSLOMNTSIWYNR--SDVEAMRLLLTSAFSLATSPAFRYDLDTLR--Q 566
DB 1356 QGRSAQPLPSTKTPAFQIQGRFVSVRQCEKASGRLAAKIIPIHPKDXTAVLREYE 1415
QY 567 AVQELVSLYEBAKSAIYS--KELASLLRAGGTLAYELLPALDEVLASDSRFLGSMLEQA 625
DB 1416 ALKGLRHPHLAQLHAAYLSPRHVLITLLEL--CSGPELLPCIAE--RASYSSESVQDYLMQM 1472
QY 626 RAAA-----VSEAEADFEYQN--SRYQULTMGPEGNILDYANKQLAGLVANYTTPRWRL 677
DB 1473 LSAQVYLHNQHLHLDLASENNIITEYNLL-----KVVDLGNAGSLSQEKVLPBDKFXD 1526
QY 678 FLEALVDSVAQG--IPFQGHQFDKXVQL--EQAFVLSKORPSQPRG 721
DB 1527 YLETMAPELLBQGAVP-----QTDIMAIGVTAFIMLSAEYVSSSEG 1568

```

RESULT 12

US-10-307-019-6  
; Sequence 6, Application US/10307019  
; Publication No. US2003010853A1  
; GENERAL INFORMATION:  
; APPLICANT: zeng, Wenlin  
; APPLICANT: Stanton, Lawrence  
; APPLICANT: SCIOS, INC.  
; TITLE OF INVENTION: MEDIATORS OF SIGNAL TRANSDUCTION  
; FILE REFERENCE: SCIOS.021DVL  
; CURRENT APPLICATION NUMBER: US/10/307,019  
; PRIOR FILING DATE: 2002-11-26  
; PRIOR APPLICATION NUMBER: 09/548,473  
; PRIOR FILING DATE: 2000-04-13  
; PRIOR APPLICATION NUMBER: 60/129,552  
; PRIOR FILING DATE: 1999-04-16  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; LENGTH: 2596  
; SEQ ID NO 6  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-307-019-6

Query Match 3.0%; Score 118.5; DB 15; Length 2596;  
Best Local Similarity 18.6%; Pred. No. 1.1;  
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

QY 19 GGAAGDEAREAAVRLVRL-LGPGPAD---FSVVERALAAKPGD-----TY 65  
DB 1746 GSAQASSSQVSSLRVSSQVGTPEPSPIDABGMTQAEADLSDSTPTLORPOEQATWRKF 1805  
QY 66 SLGG-GGAARVRVGTSGVAAAAGLHRYLRDFCG---CHVAMSGSQLR-----109  
DB 1806 SLGGRGVAGVAGYGTFAFGDAG-----GMLGQGPMMARIAVAVSQSEEEQEARAE 1859  
QY 110 -----LP-----RPLPAVPGELTEATPNHRYRYQNVCTQSYFVMMDMARWE 151  
DB 1860 SQSEEQEARAESPLPVQSARFVPEVGRAPTRSSP-----EPTPWE 1900  
QY 152 -----REI---DMMLNGINLAL-----AMSGEAIWQRYVAL-----182  
DB 1901 DIGQVSLVQIRLDGDAEADTISLDISEVDPAVYLMISLDYDIKYLPEFEMI FRKVKESA 1960  
QY 183 -----GLTQAEINER-----FTGP-----AFLANGRMGNLHTWDGP 213  
DB 1961 QPEPPSPMAEEELAEFPPTWMPGELGPPAGLEITESEBDVALLAEAVGKRRKSSP 2020  
QY 214 LPPSWHLKQLYLQHRVLDQKRSFGMTFVLPAPAGHV-----PEAVTRVFPQVNVTKMG 266  
DB 2021 SRSLFHPFG---RHLPLDEPAELGLRERYKASVEHISRILKGRPEGLEKGGP---RKKP 2074  
QY 267 SMGHFNCSYSCSFLAPEDPIFPIIGSLFIRELIKRGTDHLYGADTFNEMQPPSEPSY 326  
DB 2075 GLASFRLSGKSWDRAP-----TFRLREL-----SDETTV 2103  
QY 327 LAATTAAYEAMTAVDEAVVLLQGMLFQHQPOFWGPAQIRAVLAGVPRG---RLVLDLF 384  
DB 2104 LGQSVTLACQVSAQPAQAATWSKDGALE-----SSSRVLISATLKNFOLLITLVV 2155  
QY 385 AESQPVYTRTASFQGGPFIWCMLEHFGNGLFGALEAVNGGEAARLPNSNMGVTGMA 444  
DB 2156 AEDIGVYT-----CSVSNALGVTITTVLR-----KADR---PSSS-----2188  
QY 445 PEGISQNEVVYSIMABELGWRKDPDP-----LAAM---YTSFAARRGVHPDAGAA 493  
DB 2189 -----PCPDIGEVYADGVLLVWKPVESYGPVYIYQCSLEGSS 2226  
QY 494 WRLLRSVYNC-----SGEAC 509  
DB 2227 WTLTASDIFDCCVLTSLKSRGTYTFTTACVSKAGMGPYSSPSEOVILGAPSHLAEBS 2286  
QY 510 RGNHRSILVVRPSIQWNTSITWNR-SDFEAWRLLLTSAPSLATSPAFRYDLDLTR--Q 566

DB 2287 QGRSAQPLPSTKTFAPQTOQRGRFSVVRQWEKASGRALAALIIPIHPRKDTAVLREYE 2346  
QY 567 AVQELVSLYEARSAYLS-KEIASLIRAGCVLAYELLPALDVLASDSRFLIGSLVEQA 625  
DB 2347 ALKGLRHPHQAQHAALVSPRHVLIETL---CSGPELPLCLAE-RASYSSEVKDYLMQW 2403  
QY 626 RAAA-----VSEADAEYQON---SRYQULTMPEGNIDYANKQLAGVANYTTPRRL 677  
DB 2404 LSAQYVLANGHILHLDRSENMITEYNLL-----KVVDLGAQSLSCQKVLPSDKFKD 2457  
QY 678 FLEALYDSVAQG---IPQOHQPDFKXVFQI-EQAFVLSKORYSQPRG 721  
DB 2458 YLETMABELLEGQAVP-----QTDIWAIGVAFITLSAEYVSSBG 2499

RESULT 13  
US-10-077-130-2  
; Sequence 2, Application US/10077130  
; Publication No. US20020168742A1  
; GENERAL INFORMATION:  
; APPLICANT: Acton, Susan L.  
; APPLICANT: Kapeller-Libermann, Rosana  
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family  
; FILE REFERENCE: MPI2001-047PIKCP1(W)  
; CURRENT APPLICATION NUMBER: US/10/077,130  
; PRIOR FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: 60/269201  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; LENGTH: 2630  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-077-130-2

Query Match 3.0%; Score 118.5; DB 14; Length 2630;  
Best Local Similarity 18.6%; Pred. No. 1.1;  
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

QY 19 GGAAGDEAREAAVRLVRL-LGPGPAD---FSVVERALAAKPGD-----TY 65  
DB 1780 GSAQASSSQVSSLRVSSQVGTPEPSPIDABGMTQAEADLSDSTPTLORPOEQATWRKF 1839  
QY 66 SLGG-GGAARVRVGTSGVAAAAGLHRYLRDFCG---CHVAMSGSQLR-----109  
DB 1840 SLGGRGVAGVAGYGTFAFGDAG-----GMLGQGPMMARIAVAVSQSEEEQEARAE 1893  
QY 110 -----LP-----RPLPAVPGELTEATPNHRYRYQNVCTQSYFVMMDMARWE 151  
DB 1894 SQSEEQEARAESPLPVQSARFVPEVGRAPTRSSP-----EPTPWE 1934  
QY 152 -----REI---DMMLNGINLAL-----AMSGEAIWQRYVAL-----182  
DB 1935 DIGQVSLVQIRLDGDAEADTISLDISEVDPAVYLMISLDYDIKYLPEFEMI FRKVKESA 1994  
QY 183 -----GLTQAEINER-----FTGP-----AFLANGRMGNLHTWDGP 213  
DB 1995 QPEPPSPMAEEELAEFPPTWMPGELGPPAGLEITESEBDVALLAEAVGKRRKSSP 2054  
QY 214 LPPSWHLKQLYLQHRVLDQKRSFGMTFVLPAPAGHV-----PEAVTRVFPQVNVTKMG 266  
DB 2055 SRSLFHPFG---RHLPLDEPAELGLRERYKASVEHISRILKGRPEGLEKGGP---RKKP 2108  
QY 267 SMGHFNCSYSCSFLAPEDPIFPIIGSLFIRELIKRGTDHLYGADTFNEMQPPSEPSY 326  
DB 2109 GLASFRLSGKSWDRAP-----TFRLREL-----SDETTV 2137  
QY 327 LAATTAAYEAMTAVDEAVVLLQGMLFQHQPOFWGPAQIRAVLAGVPRG---RLVLDLF 384  
DB 2138 LGQSVTLACQVSAQPAQAATWSKDGALE-----SSSRVLISATLKNFOLLITLVV 2189

```

QY 385 AESQVYTRTASFOGQPFIMCMLHFGNHLFGALEAVNGPEARLFPNSTWVGIMA 444
| | | | |
Db 2190 AEDGVIT-----CSVNALGTVTTLVLR-----KAER--PSSS----- 2222
QY 445 PEGISQNEVYVSLMAELGMRDVPVD-----LAAM--VTSPARRYGVSHPDAGAA 493
| | | | |
Db 2223 -----PCPDIGEYVADVLLWVKPEVSEYGVITYVQCSLEGGS 2260
QY 494 WRLLRSYVNC-----SSEAC 509
| | | | |
Db 2261 WTTLASDIFDCCYITSKLSRGTYTFRACVSKAGMPYSSPSEQVLLGGPSHLASEES 2320
QY 510 RGNHNSPLVRRPSLOMTISWYNR--SDVFEAMRLLTSPSLATSPAFRYDLTLTR--Q 566
| | | | |
Db 2321 QGRSAQPLPSTKTFPAFCQIQGRGFSVVRQCKEASGALAAKIIIPHPKXTALRYE 2380
QY 567 AVQELVSLYEBARSAIYS--KELASLRAAGVLAELLPALDEVLASRFLGSMLEQA 625
| | | | |
Db 2381 ALKGLRHPHLAQHLAAVLSPEHLVLIHEL--CSGPELLPCLAE--RASYSSEVXDYLWQM 2437
QY 626 RAAA-----VSEADAFYEON--SRVQTLMGPEGNILDYANKOLAGLVANYTTPRWRL 677
| | | | |
Db 2438 LSAQVYLNQHIHLDLRSEMMITTEYNLL-----KVVDLGNQSLSQEKVLPBDKFXD 2491
QY 678 FLEALVDSVAQG--IPFOHQFDKNVFOU--EQAFVLSKORYPSQPRG 721
| | | | |
Db 2492 YLETMAPELLGQGAVP-----QTDIWAIGVAFIMLSAEVYVSSEG 2533

RESULT 14
US-10-077-130-5
; Sequence 5, Application US/10077130
; Publication No. US20020168742A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: 59079 and 12599, Protein Kinase Family
; TITLE OF INVENTION: Members and Uses Therefor
; FILE REFERENCE: MFI2001-047P1RCP1(M)
; CURRENT APPLICATION NUMBER: US/10/077,130
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269201
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 7968
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-077-130-5

Query Match 3.0%; Score 118.5; DB 14; Length 7968;
Best Local Similarity 18.6%; Pred. No. 5.8;
Matches 165; Conservative 97; Mismatches 307; Indels 319; Gaps 40;

QY 19 GGAAGDEBARAAAVRALVARL-LGPGPAD--FSVSEVERALAAKGLD-----TY 65
| | | | |
Db 7118 GSAASQASSQVSSLSRVSQVGTREPSLDAEGTQAEHLDSSTPLTQCPQEQATMKRF 7177
QY 66 SLGS-GGAARVRVNGSTGVAAAAGLHRYLRDPG--CHVAMSGQLR----- 109
| | | | |
Db 7178 SLGGRGVAGVAGYGTAFAGDAG--GMLGQGMARIMAAVQSSEEBQEBARAE 7231
QY 110 -----LP--RPLPAVPGELTEATPNNRYRYQNVQCSYFVWMDARWE 151
| | | | |
Db 7232 SQSEEQEABAESELPQVSARLPVEVGRATPSSP-----EFTPE 7272
QY 152 -----RET--DMALINGNAL-----AMSGEALWQRYVAL----- 182
| | | | |
Db 7273 DIGQVSLVQIRDLSDAEADTISLDSIEVDPAVILNLSIDYIKYLPFEEMIPRKYPKSA 7332
QY 183 -----GLTQAEINEF-----FTGP-----AFLAMGMMGLHTWDDG 213
| | | | |

```

```

Db 7333 QDEPPSPMAEELAFPEPTMWPBGELGPHAGLETTESEVDYDALLAEAVGRKKWSP 7392
QY 214 LPPSWHIKQLYQHRVLDQMSFGKTPVLPAPAGV-----PAVTRVFPQVNVTKG 266
| | | | |
Db 7393 SRSLPFFPG--RHLPDEPAELGLRERVKASVEHISILKGRPGLKEGEP--RKXP 7446
QY 267 SMGHFNCSYSGSFLAPDEPPIPIIGSLFLRELIAEFGTDHYGADTNEMQPSSEPSY 326
| | | | |
Db 7447 GLASFRLSGLSKNDAP-----TFREL-----SDEIVY 7475
QY 327 LAAATTAVEANTAVDTEAVWLQGMFLFOHQPOFNGPAQIRAVIGAVPRG--RLIVDLF 384
| | | | |
Db 7476 LQGSYTLAQVSAQPAQATWSKDGAPLE-----SSSRVLSATLKNFOLLITLVV 7527
QY 385 AESQVYTRTASFOGQPFIMCMLHFGNHLFGALEAVNGPEARLFPNSTWVGIMA 444
| | | | |
Db 7528 AEDGVIT-----CSVNALGTVTTLVLR-----KAER--PSSS----- 7560
QY 445 PEGISQNEVYVSLMAELGMRDVPVD-----LAAM--VTSPARRYGVSHPDAGAA 493
| | | | |
Db 7561 -----PCPDIGEYVADVLLWVKPEVSEYGVITYVQCSLEGGS 7598
QY 494 WRLLRSYVNC-----SSEAC 509
| | | | |
Db 7599 WTTLASDIFDCCYITSKLSRGTYTFRACVSKAGMPYSSPSEQVLLGGPSHLASEES 7658
QY 510 RGNHNSPLVRRPSLOMTISWYNR--SDVFEAMRLLTSPSLATSPAFRYDLTLTR--Q 566
| | | | |
Db 7659 QGRSAQPLPSTKTFPAFCQIQGRGFSVVRQCKEASGALAAKIIIPHPKXTALRYE 7718
QY 567 AVQELVSLYEBARSAIYS--KELASLRAAGVLAELLPALDEVLASRFLGSMLEQA 625
| | | | |
Db 7719 ALKGLRHPHLAQHLAAVLSPEHLVLIHEL--CSGPELLPCLAE--RASYSSEVXDYLWQM 7775
QY 626 RAAA-----VSEADAFYEON--SRVQTLMGPEGNILDYANKOLAGLVANYTTPRWRL 677
| | | | |
Db 7776 LSAQVYLNQHIHLDLRSEMMITTEYNLL-----KVVDLGNQSLSQEKVLPBDKFXD 7829
QY 678 FLEALVDSVAQG--IPFOHQFDKNVFOU--EQAFVLSKORYPSQPRG 721
| | | | |
Db 7830 YLETMAPELLGQGAVP-----QTDIWAIGVAFIMLSAEVYVSSEG 7871

RESULT 15
US-09-820-788-4
; Sequence 4, Application US/09820788
; Publication No. US20030166182A1
; GENERAL INFORMATION:
; APPLICANT: SHAO, Mei et al
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
; FILE REFERENCE: CL001207
; CURRENT APPLICATION NUMBER: US/09/820,788
; PRIOR FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Human
US-09-820-788-4

Query Match 3.0%; Score 117.5; DB 12; Length 497;
Best Local Similarity 21.7%; Pred. No. 0.11; Indels 153; Gaps 22;
Matches 97; Conservative 53; Mismatches 144;

QY 345 AWMLL-----QGMFLQHQPOFNGPAQIRAVIGAVRGGLVLDLFAESQPYTTRTAS 396
| | | | |
Db 14 AIFLLVDLMRRQGMARYPF--GFLPLPGL-----GNLLHYD----- 50
| | | | |
QY 397 FQGGPFIMCMLHNGN-----HGLFGLAE--VNGPEARL--FPNSTWV 439
| | | | |

```

```
Db 51 FONTPCFDOLRRRFGDVSLQAMTPVVVINGLAAREALVTHGEDTADRPVPITQIL 110
QY 440 GTGMAPEGI-----SONEVVYSIMAEIGWRDPPVDLAAMVTSFAARYGVSHD 489
Db 111 GFGPSQGVFLARYGPAMEQRFRFSVSTIRANGLOK---KSLBQWTEBA----- 158
QY 490 AGAAMRLLIRSVYNGSGEACRGHNRSPLVRPSLOMNTSWMYNSDVFEAMRLLITSAPS 549
Db 159 -----CLCAAFANHSGRPF--RPNGLDKAV---SNVI-----AS 188
QY 550 IATSPAERYD-----LIDLTOAQVOELVSIYEERASAYLSKELASILRAGVLAWE-- 601
Db 189 LTCGRREFYDDPRFLRLDLAQEGKESGFLREVINAVPVLHLI PAL--AGKVLRFOKA 246
QY 602 LLEPALDEVL-----ASDSRFLIGSWLEQARAANAVERSEADFEQNSRYQTLTNGPE 652
Db 247 FLTQLELLTEHMTWDPAQPPRDLTEAFLEAMEKAK--GNPESSFNDENLR----- 296
QY 653 GNILDYANKQAGLVANYTTPRMRLEALVDSVAQGIFFQOHQFDKXVFOLEQ----- 706
Db 297 ---IVVADLFSAGWTTSTTLAMGLMLHPDVQRV---QOEIDVIGQVREPMGDQ 350
QY 707 -----AFVLSKORYPSQPRGDTVDL 726
Db 351 AHMPYTTAVIHEVQRF-----GDIYPL 372
```

Search completed: February 13, 2004, 16:19:09  
Job time : 63 secs